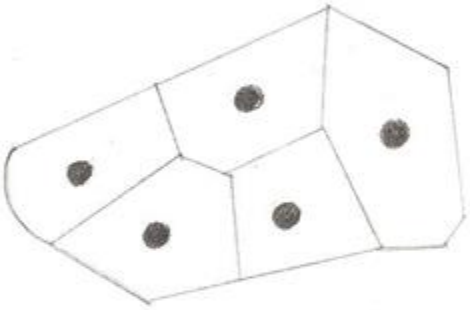
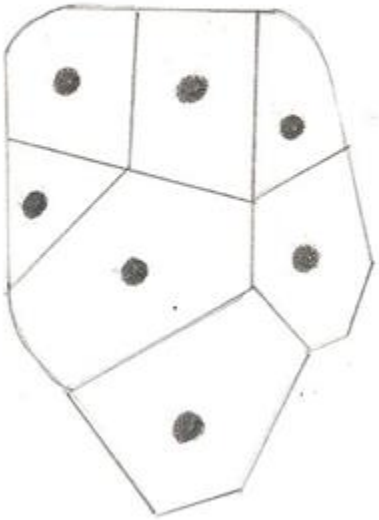
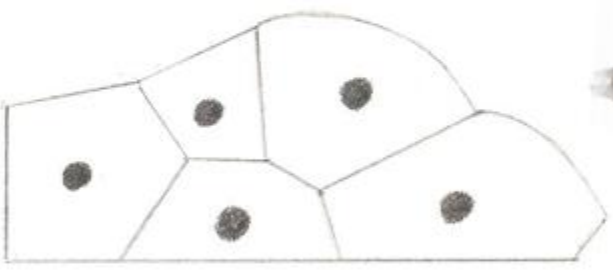


Individual variation in behaviour



Adaptive mechanisms and Ecological drivers



Alexia Mouchet

Individual variation in behaviour: adaptive mechanisms and ecological drivers

Layout: Alexia Mouchet

Illustrations: Marie-Dominique Mouchet

Individual variation in behaviour: adaptive mechanisms and ecological drivers

Dissertation
Fakultät für Biologie
Ludwig-Maximilians-Universität München

Durchgeführt am
Ludwig-Maximilians-Universität München

Vorgelegt von
Alexia Mouchet
02 August 2021

Diese Dissertation wurde angefertigt
unter der Leitung von Prof. Dr. **Niels Dingemanse**
in Bereich von **Verhaltensökologie**

Erstgutachter: Prof. Dr. **Niels Dingemanse**
Zweitgutachter: Prof. Dr. **Wolfgang Goymann**

Tag der Abgabe: **02.08.2021**
Tag der mündlichen Prüfung: **02.11.2021**

Table of contents

Summary

Wild animals show remarkable phenotypic variation despite natural selection eroding it. Phenotypic variation within populations is intriguing because all individuals are expected to be adapted to the same environmental conditions, and thus, to present similar phenotypic traits. However, when repeatedly measured, individuals have been observed to differ in the average expression of various behaviours across time and contexts. Consistent among-individual variation (called "animal personality") has been proposed to be adaptively maintained if the fitness costs and benefits of behaviour vary with the environment or other phenotypic traits. Theory postulates that two key adaptive mechanisms could play a role: life-history trade-offs and spatiotemporal variation in selection (or heterogeneous selection). Empirical tests of the role of these mechanisms in the maintenance of individual variation in behaviour remain scarce and findings are ambivalent. My PhD thesis aimed at shedding light on the mechanisms allowing the persistence of animal personalities, thereby advancing our understanding of how animals adapt to variable environments. I investigated the role of life-history trade-offs and heterogeneous selection in the coexistence of alternative personalities in the wild. I also examined potential ecological drivers of heterogeneous selection. I used a passerine bird breeding in the wild in nest boxes (the great tit *Parus major*) as model.

Individuals must trade-off investment among various phenotypic traits because they have limited amount of energy and time to acquire resources, grow and reproduce. The optimal resolution of trade-offs may depend on ecological conditions and/or the phenotypic traits of the individuals. Individuals differing in their behavioural phenotypes may thus resolve trade-offs differently. In Chapter 1, my colleagues and I tested this hypothesis by focusing on the trade-off between current reproduction and reproductive senescence. Specifically, we asked whether behavioural phenotypes differed in patterns of senescence. We found that faster explorers increased and subsequently decreased their reproductive investment with age. This finding suggests that faster explorers reproductively senesced later in life. By contrast, slower explorers laid similar clutch sizes through their lifetime; that is, they did not show reproductive senescence. Different behavioural phenotypes, thus, resolved the trade-off between current reproduction and reproductive senescence differently, which may allow them to coexist.

Spatial and temporal variation in the environment may cause natural selection to favour different phenotypes in different environments. Spatial variation in selection may maintain phenotypic variation across environments, whereas temporal variation in selection (or fluctuating selection) may maintain phenotypic variation within environments. Though these processes co-occur and may have counteracting effects on phenotypic variation, both processes have rarely been investigated simultaneously. The relative importance of spatial and temporal

variation in selection, and thus, the evolutionary potential of phenotypic traits under heterogeneous selection, remains unexplored. In Chapter 2, I studied heterogeneous selection on behaviour within and among great tit populations. To this aim, I gathered longitudinal data from five West European wild great tit populations breeding in nest boxes. In all these populations, behaviour was assayed with the same experimental design. Selection on behaviour varied primarily spatially. Temporal variation in selection was also important. The existence of phenotypic variation in all populations suggests that temporal variation played a key role in counteracting local adaption promoted by spatial variation. Temporal variation in selection was population-specific, which suggests that local ecological conditions also played a role in the evolution of phenotypic variation. This study thereby demonstrated the importance of considering both large- and small-scale geographical and temporal variation to understand the ecological mechanisms maintaining variation in animal behaviour. Previous studies found that variation in the social environment induced by variation in population density caused selection on behaviour to vary. However, we did not find such evidence in great tit populations.

Another ecological factor that varies ubiquitously and that is crucial for survival and reproduction is food availability. Food availability also generally positively correlates with population density. Therefore, the effects of population density on fitness may be indirect through food availability. Variation in food availability may cause selection pressures on behaviour to vary because behavioural phenotypes differ in competitive abilities and foraging tactics. In Chapter 3, I studied whether winter food availability drove heterogeneous selection on activity in a novel environment. I experimentally manipulated food abundance outside the breeding season by providing supplementary food in multiple great tit nest box plots. Against expectations, I did not find evidence for fecundity selection on behaviour to vary with the experimental manipulation of food availability. Food availability may drive variation in fecundity selection but simultaneous changes in breeding density may counteract its action. Food- and density-dependent selection on behaviour need to be estimated simultaneously to disentangle their effects. Interestingly, on average, individuals were more active in high than in low food availability context. Moreover, high food availability context increased behavioural variation among individuals. These findings suggest greater plasticity and/or higher survival, recruitment or immigration rate of more active individuals. Future studies should investigate whether viability rather than fecundity selection vary with food availability.

In the different projects of this PhD work, I focused on behaviour scored in different “novel environments”, which are all generally labelled “exploration behaviour”. However, “exploration behaviour” was not assayed with the same experimental design in Chapter 2

compared to Chapter 1 and 3. In Chapter 1 and 3, behaviour was assayed in the field in a portable cage. In Chapter 2, behaviour was assayed in a standardized laboratory room. We assumed that birds expressed the same behaviour in both assays because laboratory- and field-based behaviours have been shown to each correlate with other field-based behaviours. In Chapter 4, I tested this assumption and found that laboratory- and field-based behaviour did not correlate. Both assays may present different contexts to the birds, which elicited the expression of different behaviours. I also showed that the population sampled for the laboratory test was biased toward fast explorers. This study highlights the difficulty assaying behaviour in an unbiased and reproducible manner. It is therefore important to cross-validate behavioural assays before making biological assumptions.

Overall, this PhD thesis contributed to understanding the role of adaptive mechanisms in individual variation in behaviour and their ecological drivers. This work showed that behavioural phenotypes contribute differently to population dynamics and should thus be considered in ecological and evolutionary studies. This work also exemplified the importance of long-term and collaborative projects. For a comprehensive understanding of phenotypic variation, the next challenge would be to simultaneously consider multiple traits, ecological factors and species that all interact through eco-evolutionary dynamics. Such integrative studies will embrace the complexity of ecological interactions and allow us to better understand how populations adapt to variable environments.

General Introduction

INTRODUCTION

Wild animals show remarkable variation in various phenotypic traits, such as morphology, colour ornamentation, physiology or behaviour (Darwin 1859). Variation exists both among species, among populations within species, and within populations (Wilson 1998). Phenotypic variation is expected to be eroded through the process of natural selection, whereby particular traits are more likely to be passed on to the next generations and spread in the population (Darwin 1859). These traits are assumed to be traits best adapted to the environmental conditions in which the organisms live. How phenotypic variation is maintained or generated is a major question in evolutionary and behavioural ecology that remains unresolved, particularly at the within-population level.

Phenotypic variation among populations has been explained by populations inhabiting habitats with different ecological conditions causing a geographic structure of genetic mechanisms and selection pressures (Felsenstein 1976; Hoekstra 1978; Powell & Taylor 1979; Hedrick 1986, 2006; Frank & Slatkin 1990; Foster & Endler 1999). By contrast, within populations, all individuals are expected to face similar average environmental conditions, and thus, to present similar phenotypic traits adapted to their environment. The existence of phenotypic variation within populations suggests that micro-spatial variation in the environment may also play a role. Moreover, while some traits are fixed within an individual (e.g., bone length) others can change throughout lifetime, such as some morphological traits, and physiological and behavioural traits. This within-individual change in trait expression is called phenotypic plasticity.

Phenotypic plasticity can be irreversible when traits can change only once, or reversible when traits can be repeatedly expressed throughout lifetime with different levels of expression (called labile traits) (Lynch & Walsh 1998). Reversible plasticity allows individuals to respond to changes in the environment to maximise their survival and reproductive success. Reversible plasticity would be expected to be unlimited for individuals can adopt the best phenotype to the conditions. However, plasticity might be costly as it requires regularly collecting information about the environment (DeWitt *et al.* 1998). Moreover, the benefits of plasticity might be limited because the environment can change quickly, and the information collected be unreliable (DeWitt *et al.* 1998). Plasticity may thus be limited, that is, an individual may not be able to express the phenotype best fitting at any time.

Individual phenotypic variation within populations thus could result from differences in average trait expression among individuals and differences in plasticity within individuals

(Wilson 1998; Gosling 2001; Dingemanse *et al.* 2010). Individual phenotypic variation is proposed to be shaped by two types of mechanisms. Neutral mechanisms, which only involve genetic processes, and adaptive mechanisms, which involve natural (or sexual) selection (Frank & Slatkin 1990; Dingemanse & Réale 2005; Dingemanse & Wolf 2010). In this PhD thesis, I investigated the role of a number of key adaptive mechanisms proposed to maintain phenotypic variation among-individuals.

Neutral mechanisms involved in individual phenotypic variation

Most phenotypic traits are underpinned by genes, and some of the genes underpinning heritable phenotypic traits will be passed on to the offspring. The allelic frequency of these genes can be altered through generations by different genetic processes, such as genetic drift or mutations (Wright 1931; Simpson 1953; Lande 1975, 1976). Genetic drift is a change in allelic frequency caused by random mortality of individuals. In small populations, genetic drift can decrease genetic variation if some alleles disappear through generations, but in large populations, genetic drift is more likely to be small and thus maintain genetic variation stable. Mutations are alterations of the genetic sequence of a gene, which may result in a change in its function. Mutations are passed on to the offspring when they are heritable and, through the action of natural selection across several generations, spread in the population if advantageous or disappear if deleterious (Simpson 1953). When advantageous, mutations can therefore generate phenotypic variation at various levels, including among individuals within populations. Genetic pleiotropy is a third neutral mechanism that could maintain phenotypic diversity. Genetic pleiotropy occurs when a gene codes for multiple phenotypes, thereby potentially generating phenotypic diversity with limited genetic diversity. These genetic mechanisms can thus maintain or generate phenotypic variation. By contrast, natural selection is assumed to erode variation by favouring the genes best adapted to environmental conditions (Darwin 1859). The existence of phenotypic and genetic variation despite the process of natural selection suggests that neutral mechanisms and natural selection may balance each other (Lande 1975; Santiago 1998). However, several studies failed to explain the level of genetic variation observed in wild populations by mutation-selection-drift balance alone (e.g. Caballero and Keightley 1994). Another type of mechanisms must play a role in individual phenotypic variation in addition to neutral mechanisms. These mechanisms are proposed to be adaptive mechanisms.

Adaptive mechanisms involved in individual phenotypic variation

Adaptive mechanisms are mechanisms involving the action of natural (or sexual) selection as the main process shaping phenotypic variation. Three adaptive mechanisms have been proposed to play a role in individual phenotypic variation: frequency-dependent selection, life-history trade-offs and heterogeneous selection (Lewontin 1958; Clarke 1964; Levins 1969; Gillespie 1973; Stearns 1989; Frank & Slatkin 1990; Dingemanse & Wolf 2010; Wolf & McNamara 2012). These mechanisms are based on the response of natural selection to changes in the (a)biotic environment.

Frequency-dependent selection

Individuals within a population use different developmental, reproductive or behavioural strategies to survive and breed. The fitness benefits (i.e., higher reproductive success and survival) of a strategy may not only depend on the individual's own strategy, but also on the frequency of each strategy used in the population (Ayala & Campbell 1974; Maynard Smith 1974; Roff 1998). Specifically, positive frequency-dependent selection favours the common strategy, while negative-frequency dependent selection favours the rare strategy (Ayala & Campbell 1974). Positive frequency-dependent selection would thus fix the common strategy. Negative frequency-dependent selection instead could maintain multiple strategies within the population and do so in two ways (Maynard Smith 1974; Figure 1a). Selection may alternately favour, conditional on their frequency, different genotypes displaying each a fixed strategy (i.e. phenotype). Specifically, selection would favour the rare strategy until it spreads and becomes common in the population. At that point, the favoured strategy becomes disadvantageous, while the initial common disadvantaged strategy is now rare and becomes advantageous. Alternatively, selection may favour a single genotype displaying multiple strategies (i.e. plastic phenotypes). The most beneficial strategy to use would be the one the least frequently used within the population. Negative-frequency dependence may occur in various contexts, such as predator-prey interactions, when the rare prey strategy may be less conspicuous, resource competition, when using the rare strategy may allow exploiting different resources than most individuals of the population (Fitzpatrick *et al.* 2007; Calsbeek *et al.* 2010; Wolf & Weissing 2010; Mathot *et al.* 2011; Wolf & McNamara 2012) or mating interactions, when females of the rare phenotype may avoid male harassment (Svensson *et al.* 2005). Negative-frequency dependent selection, thus, could maintain either both genetic and phenotypic variation (case of one genotype displaying a fixed strategy) or only phenotypic variation (case of plastic phenotypes).

Heterogeneous selection

The biotic and abiotic environment in which organisms live varies in time and space. Spatial variation in the environment may cause natural selection to favour different phenotypes in different environments (Levene 1953; Gillespie 1974; Felsenstein 1976; Siepielski *et al.* 2013; Figure 1b). If selection pressures differ among environments, local adaptation may occur, and populations or groups of individuals within populations may diverge phenotypically and/or genetically (Grant & Price 1981; Foster & Endler 1999). Population divergence may be facilitated if gene flow among habitats is nonrandom, that is, if individuals disperse nonrandomly by settling in environments matching best their phenotypes (Hedrick 1986; Slatkin 1987; Garant *et al.* 2005; Edelaar *et al.* 2008; Clobert *et al.* 2009; Edelaar & Bolnick 2012; Nicolaus & Edelaar 2018). Population divergence may be counteracted instead if individuals do not select specific habitats (Hedrick 1986) or do not match their habitat efficiently, due to incomplete information of the environment or dispersal constraints (Slatkin 1985; Bowler & Benton 2005; Clobert *et al.* 2009). Population divergence may be counteracted also by temporal variation in ecological conditions within environments. Indeed, if the conditions change over time, the optimal phenotype to express may also change temporally within the same environment (Lande 1976). Selection pressures may thus fluctuate in direction, thereby alternately favouring different phenotypes across time (Haldane & Jayakar 1963; Hedrick 1976; Byers 2005). Consequently, the average long-term directional selection may be weak and the erosion of genetic variation within populations be low (Estes & Arnold 2007; Bell 2010; Wolf & Weissing 2010). For example, years with low food resources might induce high competition levels and thus favour relatively more competitive individuals. Years with high food resources instead might relax competition and thus favour relatively less competitive individuals due to costs of competitive abilities exceeding their benefits (Groothuis & Carere 2005; Boon *et al.* 2007). Fluctuating selection may, however, maintain phenotypic variation only in age-structured populations where individuals may face different conditions over time or if heterozygotes have highest fitness over time (heterozygote advantage) (Gillespie 1973; Frank & Slatkin 1990; Ellner & Hairston, 1994; Sasaki & Ellner 1997). Thus, spatial heterogeneous selection could maintain phenotypic variation across environments, while temporal heterogeneous selection (or fluctuating selection) could maintain phenotypic variation within environments (Hedrick 2006; Svardal *et al.* 2015). Heterogeneous selection on morphological and life-history traits has been well studied and is generally supported (Siepielski *et al.* 2009, 2013). By contrast, heterogeneous selection on behaviour has seldom been estimated (Siepielski *et al.* 2009, 2013). Therefore, it is unclear whether behavioural traits

are generally under heterogeneous selection. Though heterogeneous selection is composed of both spatial and temporal variation in selection, both processes have rarely been estimated simultaneously (Siepielski *et al.* 2013). However, both processes occur concurrently and, as explained above, may have counteracting effects on local adaptation. The relative importance of spatial and temporal variation in selection, and thus, the evolutionary potential of phenotypic traits under heterogeneous selection, remains unexplored.

Life-history trade-offs

Individuals must trade-off investment among various morphological, physiological and behavioural traits because they have limited amount of energy and time to acquire resources, grow and reproduce (Stearns 1989). A classic example of life-history trade-off in evolutionary ecology theory is the trade-off between current and future reproduction (Bell 1980; Reznick 1985; Stearns 1989). An individual could either greatly invest into current reproduction, at the expense of survival and future reproduction, or invest less into current reproduction but survive and reproduce longer. Individuals might invest more into current (vs. future) reproduction when their future reproductive expectations are low (vs. high).

The optimal trade-off resolution might vary with permanent intrinsic differences among individuals caused by differences in early life experiences (e.g., growth rate) (Stearns 1989). For example, the morphology, physiology and behaviour of an individual might affect its competitive abilities, which in turn might affect its ability to acquire the resources needed to invest more greatly into current reproduction (Réale *et al.* 2010b; Montiglio *et al.* 2018). The optimal trade-off resolution may also vary with the environment (Sih *et al.* 2004b; Stamps 2007; Wolf *et al.* 2007). Macro- or micro-spatial variation in biotic or abiotic factors might modulate the benefits of allocating into reproduction. For example, high predation risk environments might favour individuals investing more into current reproduction because of great danger of mortality, but low predation risk environments instead might favour individuals investing into future reproduction (Abbey-Lee & Dingemanse 2019). Macro-spatial environmental variation could maintain phenotypic variation in the resolution of trade-offs among populations, while micro-spatial environmental variation could maintain phenotypic variation within populations. Individuals living in different environments may thus resolve trade-offs differently while achieving similar fitness in their respective environment. Temporal variation in the environment may also allow the coexistence of multiple phenotypes within populations by alternately favouring different trade-off resolutions. Consequently, individuals

resolving trade-offs differently might achieve different fitness at any specific time point but achieve similar fitness through their lifetime and thus coexist (Figure 1c).

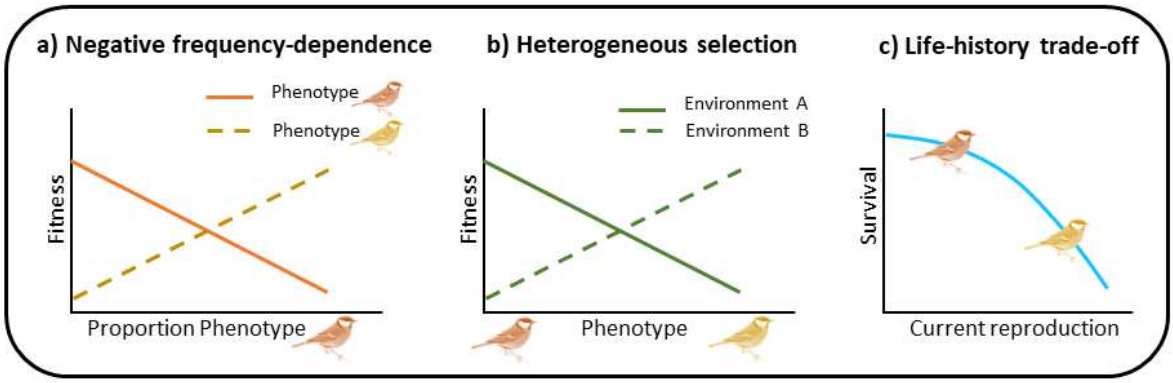


Figure 1. Illustration of key adaptive mechanisms proposed to play a role in the maintenance of phenotypic variation. a) Negative-frequency dependent selection favours phenotypes with lower frequencies in proportion to other phenotypes in the population. Illustrated is an example with two phenotypes, where phenotype Orange has highest fitness when its frequency in proportion to phenotype Yellow is low. The same applies for phenotype Yellow. b) Heterogeneous selection favours different phenotypes in different environments, with phenotype Orange having highest fitness in Environment A (straight line) and phenotype Yellow in Environment B (dashed line). c) Life-history trade-offs may allow different phenotypes to coexist within populations if the phenotypes resolve trade-offs differently, for example, by investing less into current reproduction in favour of survival (phenotype Orange) or investing more into current reproduction at the expense of survival (phenotype Yellow)

The specific case of behaviour

Behaviour is a highly reversibly plastic type of phenotypic trait. Thus, all individuals within a population would be expected to be able to behave similarly in a same context. However, behaviour varies among individuals within populations. If plasticity is limited, among-individual variation in behaviour may be explained by individual variation in plasticity levels (Dingemanse *et al.* 2010). Variation in plasticity does not explain, however, the observation of individuals consistently differing in their average behaviour that have been described in many taxa (Wilson 1998; Gosling 2001; Sih *et al.* 2004a; Réale *et al.* 2007). Indeed, when repeatedly measured, some individuals express higher average levels of behaviour than others across contexts and time. These differences have been called repeatable individual differences or animal personality (Wilson 1998). How animal personalities coexist within populations is a major question in behavioural ecology because it implies that multiple behavioural phenotypes achieve similar fitness despite the action of natural selection (Sih *et al.* 2004a; Dingemanse & Réale 2005; Réale *et al.* 2007).

Animal personalities have been proposed to adaptively coexist if the fitness costs and benefits of behaviour varies with the environment or other phenotypic traits, such as morphology or physiology (Dingemanse and Réale 2005, 2013; Dingemanse and Wolf 2010; Réale et al. 2010; Wolf and Weissing 2010). In other words, personalities may coexist if individuals differ in their reproductive assets and the risks they are willing to take in losing these assets (Wolf *et al.* 2007; Luttbeg & Sih 2010; Wolf & Weissing 2010). Individuals consistently differing in behaviours that may allow higher resources acquisition at the expense of survival, called risk-taking behaviours, have been shown to differ in life-history traits (Biro & Stamps 2009), dispersal abilities (Dingemanse *et al.* 2003; Duckworth 2008; Cote *et al.* 2010; Quinn *et al.* 2011; Cooper *et al.* 2017), and behavioural plasticity (Coppens *et al.* 2010; Dingemanse *et al.* 2012). The adaptive mechanisms described above, frequency-dependent selection, heterogeneous selection or life-history trade-offs, thus may play a key role in the coexistence of animal personalities (Figure 2). Though these proposed adaptive mechanisms are strong and important explanations for the maintenance of individual variation in behaviour, empirical tests of their role remain scarce or findings are ambivalent (Boon et al. 2007; Cote et al. 2008; Quinn et al. 2009; Nicolaus et al. 2012, 2016; Taylor et al. 2014; Le Cœur et al. 2015; Santostefano et al. 2017; Royauté et al. 2018).

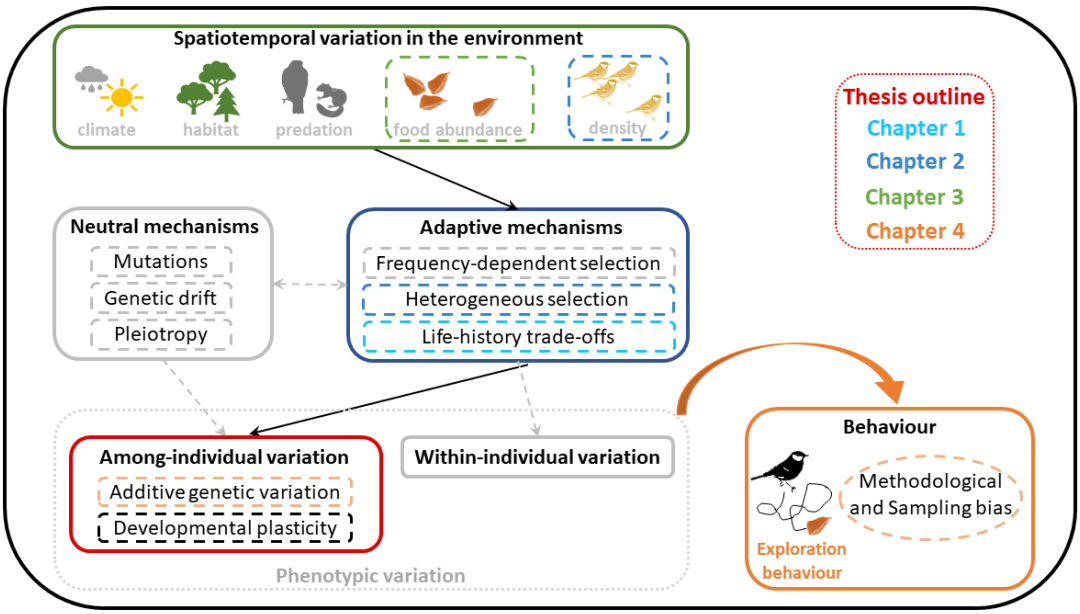


Figure 2. Overview of the framework explaining the existence of individual phenotypic variation. This thesis focused on better understanding the coexistence of repeatable among-individual variation in behaviour (i.e., animal personality; red box) using exploration behaviour as focal behaviour (orange box). This work examined the role played by two of the proposed adaptive mechanisms, heterogeneous selection and life-history trade-offs (blue box). It also studied which ecological factors could drive heterogeneous selection, focusing on population density and food availability (green box).

Thesis aims

In this thesis, I aimed to shed light on the mechanisms leading to the persistence of repeatable among-individual variation in behaviour. My thesis thereby improves our understanding of the evolutionary potential of phenotypic traits, and of how populations adapt to their environment. Specifically, I examined the role of heterogeneous selection and life-history trade-offs in the co-existence of alternative personalities. I also aimed at determining which ecological factors could induce heterogenous selection using an experimental approach. I conducted this research on long-term datasets of a passerine species, the great tit (*Parus major*), and focusing on exploration behaviour. Finally, in animal personality research, a recurring issue is how to assay behaviour in a standardized manner, while being biologically meaningful. Different methods may be used to assay a behaviour assumed to have the same ecological function in the different contexts. I tested this assumption for 'exploration behaviour' that was scored using two types of assays across my studies. After a brief explanation of why using the great tit as model and focusing on exploration behaviour to answer this research questions, I briefly describe the four chapters that compose this dissertation (Figure 1).

The great tit as model system

Great tits are an ideal system to study the adaptive maintenance of individual variation in behaviour because their life-history is well known, and personalities have been well described in multiple populations (Perrins 1965; Krebs 1971; Harvey *et al.* 1979; Balen 1980; Gosler 1993; Verbeek *et al.* 1994, 1996; Dingemanse *et al.* 2002). The great tit is a common passerine bird inhabiting deciduous and mixed deciduous woodlands in most Eurasia. Great tits breed in tree cavities and readily accept breeding in nest boxes, which facilitates their reproductive success monitoring in the wild. In Western Europe, males defend a breeding territory from January-February onwards and the pair starts breeding in April. The great tit is a socially monogamous species, but extrapair mating occurs. This is a short-lived species with an average lifespan of 2-3 years, but some individuals can live up to 10 years (Gosler 1993). Populations of great tits are thus age-structured. The female lays 7-8 eggs on average, but a clutch can range from 4 to 12 eggs (Lack 1964; Perrins 1965). After an incubation period of about 12 days, the offspring hatch and stay in the nest for 19-21 days, during which time they are fed by both their social parents (Gosler 1993). Because both parents regularly feed their offspring, they can easily be captured in their breeding nest box. At first capture, individuals are uniquely marked to allow their identification. Because breeding dispersal is rare in this species (Harvey *et al.* 1979), the same individuals can be subsequently captured, allowing the acquisition of repeated

measures of phenotypic traits. These capture events also allow us to collect behavioural data in the field, such as exploration behaviour. Great tits can be caught outside the breeding season as well, either when they roost at night in nest boxes or using mist-nets.

Exploration behaviour

Exploration behaviour is defined as the activity of an individual expressed in a novel environment. It is generally scored as the sum of movements the individual does between different areas of the novel environment (Verbeek *et al.* 1994; Dingemanse *et al.* 2002; Boon *et al.* 2007). In many taxa and species, including the great tit, exploration behaviour is repeatable and heritable (Dingemanse *et al.* 2012; Korsten *et al.* 2013; Santure *et al.* 2015; Nicolaus *et al.* 2016a); thus, this trait can be under selection and has the potential to evolve. Exploration behaviour is considered a type of risk-taking behaviour because it has been shown to correlate with other risk-taking behaviours, such as aggressiveness or boldness (Verbeek *et al.* 1994; Hollander *et al.* 2008; Amy *et al.* 2010; Mutzel *et al.* 2013; Stuber *et al.* 2013). The exploration of a novel environment may relate to a higher predation or starvation risk, given the absence of prior information in the environment. Risk-prone individuals also are predicted to have a faster pace-of-life (i.e., fast growth and early reproduction) and higher metabolism compared to risk-adverse individuals. Pace-of-life theory indeed postulates that individuals range along a slow-fast continuum of life-history, metabolism and behaviour (Réale, Garant, et al. 2010; Dammhahn et al. 2018; Montiglio et al. 2018; but see Niemelä and Dingemanse 2018; Royauté et al. 2018). Individuals differing in their behaviour may vary in their resource acquisition and ability to invest in life-history traits, and therefore, in the resolution of life-history trade-offs (Montiglio *et al.* 2018; Laskowski *et al.* 2020; Moiron *et al.* 2020). As mentioned above, fitness costs and benefits of risk-taking behaviours may vary with the environment and other phenotypic traits. Selection pressures on exploration behaviour may thus change with the risk level of the environment, favouring risk-prone individuals in safer environments and risk-adverse individuals in risky environment. Exploration behaviour is thus suitable to examine the role of adaptive mechanisms, such as heterogeneous selection and life-history trade-offs, in the coexistence of personalities within populations.

Thesis outline

In Chapter 1, my colleagues and I tested whether the life-history trade-off between current and future reproduction was resolved differently by individuals differing in their behavioural phenotype. Risk-prone individuals, i.e., faster explorers are expected to invest more into current

reproduction at the expense of future reproductive assets. Future assets may be lost if chances of survival diminish. Results of empirical studies testing the trade-off between current reproduction and survival are ambivalent. Investment into current reproduction may instead trade off with reproductive senescence. That is, individuals that highly invest into reproduction early might decrease their investment into reproduction earlier in age. Studies have rarely investigated whether such trade-off was resolved differently by different behavioural phenotypes. We therefore examined whether behavioural phenotypes differ in their investment into current reproduction and patterns of senescence. We did so while accounting for among-year plasticity in behaviour and reproduction. This is important because within-individual effects can mask among-individual effects when within- and among-individual effects show opposite patterns. We hypothesized that faster explorers would invest more into current reproduction at the expense of earlier reproductive senescence.

In chapter 2, I investigated the relative importance of spatial and temporal variation in selection to better understanding the evolutionary potential of behavioural traits among and within populations. I gathered longitudinal data from five West European great tit populations located in 4 different countries (The Netherlands, Belgium, United Kingdom and Germany). This longitudinal dataset with multiple populations and multiple study plots within populations allowed us to uniquely compare variation in selection in time and space and at macro- and micro-scales. Indeed, we simultaneously estimated spatial variation in selection among and within populations (i.e., at macro- and micro-spatial scale), temporal variation across all populations, and population-specific temporal variation. This large-scale collaborative project was possible because, in all these populations, exploration behaviour was assayed using the same experimental design. I postulated that spatial variation in selection would favour phenotypic divergence among populations and among habitats within populations. I predicted that temporal variation in selection instead would favour the coexistence of multiple behavioural phenotypes within populations. Temporal variation in selection may therefore play a key role in counteracting local adaptation and thus population divergence.

Spatiotemporal variation in selection is likely induced by spatiotemporal changes in ecological factors in the environment. However, we still poorly understand what ecological factors, biotic or abiotic, drive heterogeneous selection on behaviour. Food availability and population density are two factors known to vary drastically in most species (e.g., Lack 1954). Density-dependent selection is an important driver of population dynamics and a few studies have found that selection on behaviour varied with density (Cote *et al.* 2008; Le Galliard *et al.* 2015; Nicolaus *et al.* 2016b). In Chapter2, I tested the generality of this finding in great tit

populations. An increase in population density is often coupled with an increase in food availability (Perrins 1965; Balen 1980; Källander 1981; Perdeck *et al.* 2000; Prevedello *et al.* 2013). For example, in great tits, higher winter food availability increases breeding density by increasing survival of yearling and older adults, and likely by attracting immigrants (Perrins 1965; Balen 1980; Källander 1981; Grøtan *et al.* 2009). Therefore, observational patterns of fluctuating density-dependent selection on behaviour may partly be driven by food availability. Food resources are crucial for survival and reproduction, and behaviour modulates their access (Verbeek *et al.* 1996; Crates *et al.* 2016; Milligan *et al.* 2017; Moiron *et al.* 2018). Thus, food availability is expected to be a major driver of heterogeneous selection on behaviour. A few studies investigating the link between fitness and personalities in great tits and North American red squirrels (*Tamiasciurus hudsonicus*) observed that variation in selection coincided with variation in winter food availability (Dingemanse *et al.* 2004; Boon *et al.* 2007). In chipmunks (Tamia sibiricus), selection on behaviour correlated with tree masting (Le Cœur *et al.* 2015). However, the causal link between variation in food availability and selection has not yet been tested. In Chapter 3, I experimentally tested whether winter food availability drives heterogeneous selection on exploration behaviour in great tits. I provided supplementary food outside the breeding season in 12 nest box plots of great tits monitored south of Germany for four consecutive years. I estimated selection on exploration behaviour in high versus low food availability environments. As a previous studies found that selection favours faster explorers in low densities (Nicolaus et al. 2016), I hypothesized that faster explorers would be favoured in low food availability when density is expected to be relatively low. By contrast, I expected slower explorers to be favoured in high food availability contexts. Stressful conditions may not allow the expression of all behavioural strategies because some of them would be too costly in this context, for example in terms of survival. Favourable conditions, by contrast, may allow more behavioural strategies to be expressed because the costs of different strategies might not substantially differ. Alternatively, intraspecific competition for resources may lead individuals to diversify behavioural tactics to release the competition level. That is, individuals may specialise on different niches, for examples, by preying different items or foraging in different places, to reduce competition with other conspecifics (van Valen 1965; Bolnick *et al.* 2003; Svanbäck & Bolnick 2007; Bergmüller & Taborsky 2010). Consequently, variation in behaviour among individuals would increase. Two processes could be at play. Individuals may plastically adjust their behaviour to the context or selection may be strong against certain types of individuals in highly competitive contexts and more relaxed in less competitive contexts (Bolnick *et al.* 2007). Competition for food resources is expected to be highest when food is

scarce. However, competition is also expected to be highest when population density is high, which is often coupled with high food availability. Depending on which factor induces stronger competition for food resources, niche specialisation, and thus variation in behavioural phenotypes, may be stronger in either low or high food availability context. I hypothesized that among-individual variation in exploration behaviour would differ between food supplemented and non-supplemented conditions. Given that supplemented food is clumped and limited to a few feeders, population density may have a strong effect in such conditions compared to when food is more evenly distributed. Therefore, among-individual behavioural variation may be larger in food supplemented conditions, either due to more favourable food conditions or stronger intraspecific competition.

Behaviour is generally scored in standardized behavioural assays in the laboratory because it is often difficult to record and score in the wild. In animal personality research that aimed at quantifying behavioural differences among individuals within and across contexts, fully standardized assays deemed important. However, different environments can elicit the expression of different genes (called gene-by-environment interactions), which may result in different behaviours and changes in behavioural variation (Hoffmann & Merilä 1999; Charmantier & Garant 2005; Niemelä & Dingemanse 2014). Therefore, behaviour expressed in artificial environments may not reflect behaviour expressed in natural environments (Houle *et al.* 2011; Carter *et al.* 2013; Niemelä & Dingemanse 2014). In great tits, exploration behaviour is generally assayed in a laboratory room outside the breeding season (Verbeek *et al.* 1996; Dingemanse *et al.* 2002; Figure 3). Birds captured in their roosting nest box at night, or with mist nets, are brought to the laboratory. After an overnight stay, their exploration behaviour is assayed in a standardized room that represents a novel environment. This behaviour is assumed to reflect behaviour in the wild. However, laboratory-based behaviour has not been cross-validated in the field to ensure both laboratory- and field-based behaviour are expression of the same phenotypic trait (Carter *et al.* 2013). Moreover, this validation should be performed before comparing results of and formulating hypotheses on seemingly similar behaviours assayed using different designs as these designs may present different environments. In our great tit population in southern Germany, an assay to score exploration behaviour in the field has been developed. This assay consists in releasing a bird in a portable cage that represents a novel environment (Stuber *et al.* 2013; Figure 3). This assay initially had been conceived to avoid capture biases towards certain behavioural phenotypes outside the breeding season as certain types of birds might preferentially roost in nest boxes. The field-based assay can also be performed during the breeding season, which allows us to assay most

breeding birds. Sample sizes for studies linking fitness and behavioural data can therefore be enlarged. In this thesis, I used 'exploration behaviour' assayed in both contexts: for chapter 1 and 3, behaviour was scored in the field and for chapter 2, behaviour was scored in the laboratory. I assumed behaviour expressed in both contexts was the same trait because laboratory- and field-based behaviours both correlate with field-based aggressiveness and boldness behaviours (Stuber *et al.* 2013; Moiron *et al.* 2019). However, a direct cross-context validation had not be performed. In chapter 4, I therefore investigated whether 'exploration behaviour' assayed in the laboratory and the field represented the same behavioural trait. First, I estimated the heritability of behaviour in each context, as selection can only act on heritable traits. Second, I estimated the genetic cross-context correlation between laboratory- and field-based behaviour using quantitative genetics approaches. I predicted that laboratory- and field-based behaviour will be correlated, representing the same behavioural trait. Finally, I tested whether sampling bias with regards to behavioural phenotypes occurred in the sample assayed in the laboratory.

Figure 3. Experimental designs used to score "exploration behaviour". On the left, laboratory-based assay in a room equipped with five artificial "trees". On the right, field-based assay in a cage equipped with three perches and a front grid. Before starting the assay, the bird was placed on the side box attached to the cage.

Chapter 1

Individual variation in age-dependent reproduction: Fast explorers live fast but senesce young?

Published in Journal of Animal Ecology (2020), **89** (2), 601-613
https://doi.org/10.1111/1365-2656.13122

Individual variation in age-dependent reproduction: Fast explorers live fast but senesce young?

Niels J. Dingemanse, Maria Moiron, Yimen G. Araya-Ajoy, Alexia Mouchet, Robin N. Abbey-Lee

Abstract

1. Adaptive integration of life history and behaviour is expected to result in variation in the pace-of-life. Previous work focused on whether 'risky' phenotypes live fast but die young, but reported conflicting support. We posit that individuals exhibiting risky phenotypes may alternatively invest heavily in early-life reproduction but consequently suffer greater reproductive senescence.

2. We used a 7-year longitudinal dataset with >1,200 breeding records of >800 female great tits assayed annually for exploratory behaviour to test whether within-individual age dependency of reproduction varied with exploratory behaviour. We controlled for biasing effects of selective (dis)appearance and within-individual behavioural plasticity.

3. Slower and faster explorers produced moderate-sized clutches when young; faster explorers subsequently showed an increase in clutch size that diminished with age (with moderate support for declines when old), whereas slower explorers produced moderate-sized clutches throughout their lives. There was some evidence that the same pattern characterized annual fledgling success, if so, unpredictable environmental effects diluted personality-related differences in this downstream reproductive trait.

4. Support for age-related selective appearance was apparent, but only when failing to appreciate within-individual plasticity in reproduction and behaviour.

5. Our study identifies within-individual age-dependent reproduction, and reproductive senescence, as key components of life-history strategies that vary between individuals differing in risky behaviour. Future research should thus incorporate age-dependent reproduction in pace-of-life studies.

Keywords: age dependence, behaviour, life history, personality, reaction norms, reproduction, senescence, variance partitioning

INTRODUCTION

Life-history theory predicts that organisms resolve trade-offs between current and future reproduction differently depending on ecology (Saether, 1988; Stearns, 1992; Williams, 1966). This may cause variation in life histories along a pace-of-life (POL) slow-to fast continuum (Ricklefs & Wikelski, 2002). Comparative research demonstrated covariance ('syndrome' structure) between behavioural, physiological and life-history traits among species or populations, called a pace-of-life syndrome (POLS) (Ricklefs & Wikelski, 2002). Current POLS studies address whether among-individual behavioural differences (aka 'personality') co-evolved with POL within populations (Dammhahn, Dingemanse, Niemelä, & Reale, 2018; Réale et al., 2010). Research concentrates on 'risky behaviours' (e.g. aggressiveness, anti-predator boldness, exploration) that facilitate resource acquisition at the cost of reduced life span and may thus function as mediators of life-history trade-offs (Biro & Stamps, 2008, 2010; Careau, Thomas, Humphries, & Réale, 2008; Stamps, 2007; Wolf, Doorn, Leimar, & Weissing, 2007).

Studies of within-population POLSs imply that aggressive, bold or explorative individuals exhibit a 'fast' lifestyle characterized by fast growth, early maturation, increased reproductive output per breeding attempt and a reduced life span. Adaptive theory implies such patterns result from individual variation in residual reproductive value (reviewed by Dingemanse & Wolf, 2010): individuals with low residual reproductive values disproportionally benefit from risky behaviours because they gain reproductive benefits but lose little when such actions reduce life span (Wolf et al., 2007). Support comes from manipulations of early-life conditions, and parental effort, demonstrating that risky behaviour is up- versus down-regulated when residual reproductive value is decreased versus increased (Bateson, Brilot, Gillespie, Monaghan, & Nettle, 2015; Nicolaus et al., 2012). Various studies have already demonstrated that bold individuals 'live fast but die young', confirming POLS-theoretical predictions (reviewed by Réale et al., 2010; Royaute, Berdal, Hickey, & Dochtermann, 2018; Smith & Blumstein, 2008). Other studies, by contrast, report zero or opposite relationships between risky behaviours, reproduction and life span (e.g. Nicolaus, Piault, Ubels, Tinbergen, & Dingemanse, 2016; Niemelä, Dingemanse, Alioravainen, Vainikka, & Kortet, 2013; Santostefano, Wilson, Niemelä, & Dingemanse, 2017); the validity of POLS concept is therefore subject to debate (Mathot & Frankenhuis, 2018; Royaute et al., 2018).

Pace-of-life syndrome studies, however, fail to appreciate that trade-offs between current and future reproduction may, depending on ecology, be resolved in multiple ways (Montiglio,

Dammhahn, Messier, & Reale, 2018). That is, POLS research has focussed on survival costs associated with fast life histories (Royaute et al., 2018; Smith & Blumstein, 2008), while the cost of reproduction can also be expressed by an earlier onset of reproductive senescence (Lemaitre et al., 2015). Reproductive senescence is the age-dependent decline in reproductive performance within individuals due to deteriorating physiological and cellular functioning when older, evolved because extrinsic mortality weakens selection with increasing age (Fisher, 1930; Hamilton, 1966; Medawar, 1952; Williams, 1957). Individuals with risky behavioural profiles (as defined above) may thus pay the costs of their fast lifestyle (increased reproductive output per breeding attempt) by reproductively senescing earlier in life. This is in line with laboratory studies showing that bold fish suffer greater oxidative stress and faster telomere attrition (Pauliny, Devlin, Johnsson, & Blomqvist, 2015), while bold fish also have shorter telomeres in the wild (Adriaenssens, Pauliny, Blomqvist, & Johnsson, 2016). The hypothesized integration of reproductive senescence as part of a POLS predicts individuality in age-dependent reproduction within populations, for which ample evidence exists (e.g. Brommer, Rattiste, & Wilson, 2010; Brommer, Wilson, & Gustafsson, 2007; Evans, Gustafsson, & Sheldon, 2011). It further predicts that fast life histories are associated with earlier reproductive senescence, as demonstrated by among-species comparisons (Jones et al., 2008).

By contrast, few studies investigated whether among-individual differences in risky behaviour covary with age-dependent reproduction (Patrick & Weimerskirch, 2015; Réale, Martin, Coltman, Poissant, & Festa-Bianchet, 2009). Importantly, associations between reproduction and age result from two distinct processes (van de Pol & Verhulst, 2006). Reproduction varies with age within individuals, first, due to age-related plasticity, and second, due to selective (dis)appearance of low- versus high-quality individuals. For example, individuals producing large clutch sizes throughout their lives ('high-quality' individuals) may also start reproducing when young, or have a long reproductive life. The hypothesized integration of risky behaviour and age-dependent reproduction posits that within-individual age-related plasticity varies among behavioural types, requiring approaches that disentangle within- from among-individual age effects (van de Pol & Verhulst, 2006). Similarly, risky behaviours differ among individuals (Bell, Hankison, & Laskowski, 2009; Holtmann, Lagisz, & Nakagawa, 2017) but simultaneously exhibit within-individual age-dependent plasticity (Araya-Ajoy & Dingemanse, 2017; Brommer & Class, 2015; Class & Brommer, 2016; Fisher, David, Tregenza, & Rodriguez-Munoz, 2015; Patrick, Charmantier, & Weimerskirch, 2013). Repeated measures are thus required to estimate relationships between individual-level behaviour and reproductive senescence while avoiding bias due to within-individual plasticity

(Niemelä & Dingemanse, 2018a, 2018b). To our knowledge, this is the first study of personality-related age dependency of reproduction that fully applies such approaches.

We tested whether individuals exhibiting risky behavioural profiles also allocated more resources to (early-life) reproduction, and whether they suffered greater reproductive senescence. We used a descriptive approach, acknowledging that experimental studies will be required to test whether personality-related allocation to early-life reproduction represents an investment causally affecting reproduction later in life. We used a longitudinal dataset with 1,209 breeding records of 813 female great tits assayed annually during the reproductive phase for their activity in a small cage (Stuber et al., 2013). Our previous studies demonstrated that activity represents a risky behaviour, implying that it allows for an appropriate test of theory (sensu Carter, Feeney, Marshall, Cowlishaw, & Heinsohn, 2013; Houle, Pelabon, Wagner, & Hansen, 2011). Active great tits—called 'faster' explorers throughout—behave more boldly when confronted with risk of predation (Stuber et al., 2013) and respond more aggressively to territorial intrusions than 'slower' (less active) explorers (Moiron, Araya-Ajoy, Mathot, Mouchet, & Dingemanse, 2019). In line with POLS predictions, faster great tits also produce larger clutches (Araya-Ajoy et al., 2016) and are more willing to shift investment towards current reproduction when given the opportunity (Nicolaus et al., 2015).

We aimed to estimate within-individual age dependency of annual reproduction, focussing on four reproductive traits determining annual reproductive success: clutch size, nest success (binary probability to produce any fledglings), and for successful nests, fledgling number and average mass. For each trait, we estimated within-individual age dependency of reproduction as a function of exploratory behaviour. Our repeated measures design enabled estimating relationships between individual-level behaviour and reproductive senescence while avoiding bias caused by within-individual plasticity (Niemelä & Dingemanse, 2018a, 2018b).

MATERIALS AND M ETHODS

Field methodology

The study was performed in 12 nest box plots in mixed deciduous forests within a 15 × 20 km2 area near Munich, Germany (47°58′N, 11°14′E). Each plot consisted of 50 boxes within a regular grid covering ~9 ha. For 7 years (2010–2016), nest boxes were inspected (bi)weekly (April–July) to record lay date (back-calculated assuming one egg laid per day) and clutch size. Shortly before expected hatching, boxes were inspected daily to determine hatch date (day 0). At day 7, each parent was captured with a spring trap inside the box, marked with an aluminium

ring and a unique colour ring combination (if not banded previously), and assayed for their activity in a cage (Stuber et al., 2013). This assay represents a version of the classic 'novel environment test' (Dingemanse et al., 2012; Verbeek, Drent, & Wiepkema, 1994) modified for field research (Kluen & Brommer, 2013; Stuber et al., 2013). Briefly, the subject's behaviour was recorded for 2 min with a camera placed 1.5 m in front of the cage (detailed in Stuber et al., 2013). The total number of hops among cage locations was used as a proxy for exploratory behaviour (Araya-Ajoy et al., 2016), where faster explorers had higher scores. Directly following testing, sex and age (first-year breeder vs. older) were determined (based on plumage characteristics; Jenni & Winkler, 1994), standard morphological measurements (body mass, tarsus, bill and wing length) and a blood sample taken, and the bird released (within 15 min post-capture). On day 9, another capture attempt was made if we previously failed to capture both parents. On day 14, mentioned morphological traits were measured for all nestlings alive. Boxes were inspected every second day from day 19 onwards to determine fledgling number. Outside the breeding season, boxes were inspected at night (once or twice per winter), and roosting individuals captured and ringed (Abbey-Lee, Mathot, & Dingemanse, 2016; Mathot, Nicolaus, Araya-Ajoy, Dingemanse, & Kempenaers, 2015; Stuber et al., 2013); the exploration test in the cage was not conducted at this time.

Statistical analyses

We first produced a base model estimating population-average within-individual age effects, and the population-average age of peak performance, for key determinants of reproductive success (n = 1,209) of 'first clutches' (clutches initiated within 30 days after the first clutch of the year was found; van Noordwijk, McCleery, & Perrins, 1995). We focused on clutch size, average offspring body mass at day 14 and number of offspring fledged. Visual inspection of raw data and residuals of models (detailed below) showed that traits were sufficiently normally distributed; however, for fledgling number this was only so when excluding first broods failing completely (n = 315 of 1,209 nests; 26%) (Appendix S1). We therefore studied variation in fledgling number by analysing, first, the binary probability to fledge any offspring (n = 1,209 nests), and, second, for successful nests, fledgling number (n = 894 nests). We chose this approach to reduce the number of distributional assumptions, and analytical complexity, associated with alternative (e.g. zero-inflated Poisson) models. Analyses of the binary probability to fledge any offspring implied that total nest failure occurred randomly with respect to key predictors; this was also the case for expanded models (detailed below) where effects of exploratory behaviour were never strongly supported (Appendix S2, Table S2). The

subset of nests producing fledglings (n = 894 of 1,209 nests; 74%) thus appeared to represent an unbiased sample; total nest failure is therefore not discussed further. Integrative measures of reproductive fitness, such as the number of offspring recruiting as breeders into the population (Bouwhuis, Sheldon, Verhulst, & Charmantier, 2009), could not be used because our study setup (small nest box plots within larger patches of suitable habitat) resulted in little local recruitment (Nicolaus et al., 2015). As a second step, we constructed an expanded model to determine whether an individual's average level of exploratory behaviour (defined below) predicted its age-dependent reproductive profile. Our previous studies showed that reproductive parameters (like clutch size) are repeatable with respect to female but not male identity (Araya-Ajoy et al., 2016). As our primary interest was in analysing effects of repeatable (i.e. among-individual) differences of exploratory behaviour, we thus focussed on female breeders throughout.

Defining age categories

We defined age in years since birth, with age = 0 representing the year of birth; great tits breed earliest as 1-year-olds (age = 1). Absolute age was known for any breeder ringed as nestling in our populations ('local recruit'; n = 77 of 813 birds, 9%). The majority of these local recruits bred as 1-year-olds (n = 69 of 77 local recruits, 90%). Absolute age could also be determined for unringed birds identified, based on plumage characteristics, as 1-year-olds (n = 529 of 736 immigrant recruits, 72%). Absolute age could not be determined for immigrants first captured with an adult plumage (implying they were 2-year-olds or older, age $\geq$ 2; n = 207 of 736 immigrant recruits, 28%). Following Bouwhuis et al. (2009), this latter category of immigrants was assumed to have recruited as 2-year-olds. Local recruits not recruiting as 1-year-olds (n = 8), all recruited as 2-yearolds, validating this assumption.

Modelling age effects

Following Bouwhuis et al. (2009), statistical analyses fitted linear and quadratic age to simultaneously model pre-peak improvements and post-peak declines in reproduction. All analyses also fitted 'first observed age' and 'last observed age' of reproduction to control, respectively, for selective appearance and disappearance from the dataset of birds differing in average annual reproductive performance; this avoids biases in estimates of within-individual age effects (van de Pol & Verhulst, 2006). First observed age of reproduction, determined using breeding season and roosting captures (see above), was 1 (n = 598 females; 74%), 2 (n = 190; 24%), 3 (n = 13; 2%), 4 (n = 3; <1%) or 5 (n = 1; <1%). Fewer than 2% of all females (n = 17 of 813 individuals) were (older than) 3 years old at first observed age of reproduction; we

therefore pragmatically fitted first observed age as a two-level factor in our analyses (recruited as 1 year old vs. older). Notably, no bird recruiting as a 3 years old or older had breeding records (e.g. second or replacement clutches) from previous years. Rather, those were immigrants previously ringed in our study area (e.g. in winter; see above), that had likely bred previously in natural cavities, whether adjacent to our study area (Drent, 1984) or elsewhere (Harvey, Greenwood, & Perrins, 1979). Last observed age of reproduction was 1 (n = 414 females; 51%), 2 (n = 237; 29%), 3 (n = 101; 12%), 4 (n = 44; 5%), 5 (n = 12; 1%), 6 (n = 4; <1%) or 7 (n = 1, <1%). Controlling for differences in last observed age effects between birds with complete life histories (defined as birds not observed for two consecutive years following their last observed productive event; Bouwhuis et al., 2009) versus incomplete life histories (all other birds) did not bias parameters of key interest (Appendix S3 and Table S3a). The same was true when controlling for female body mass (Table S3b). We therefore ignored these variables in analyses reported in the main text.

Base models

Age effects were modelled by fitting (for each trait separately) a univariate mixed-effect model, where a statistical intercept (β0), age (β1), age squared (β2), first observed age (β3) and last observed age (β4) were included as fixed effects (age variables as covariates except for first observed age, see above). Age was fitted as age-1 to ensure that intercepts of our models represented the reproductive performance for the earliest age of first reproduction. Random intercepts were included for individual, plot, year and plot-year identity (unique combination of plot and year); for sample sizes, see Table 1. The latter three random effects controlled, respectively, for unmeasured spatial, temporal and spatiotemporal environmental effects (Araya- Ajoy & Dingemanse, 2017; Araya-Ajoy et al., 2016). We further controlled for brood size manipulations conducted in 2010 and 2011 (detailed in Appendix S4). Previous analyses showed that slower explorers had highest reproductive success when given experimental brood sizes equal to their natural choice, while faster explorers had highest reproductive success when given increased brood sizes (Nicolaus et al., 2015). Neither reproductive traits (e.g. clutch size, fledgling number) nor exploratory behaviour were affected by perceived predation levels (manipulated in 2013 and 2014; see Table S1 in Abbey-Lee & Dingemanse, 2019). Exploratory behaviour also did not vary with observer identity (Moiron et al., 2019). We therefore did not consider these factors further. Models assumed a binomial (probability to produce any fledglings) or Gaussian error distribution (all other traits).

For any reproductive trait with statistical evidence (defined below) for quadratic within-individual age effects, we also estimated (a) the age of peak reproduction as $-\beta_1/2\beta_2$, and (b) the associated reproductive performance at this age ('peak performance') as $\beta_0 - \beta^2_1/ 4\beta_2$ (Bronshtein, Semendyayev, Musiol, & Mühlig, 2015); the uncertainty associated with these derived parameters was calculated by taking forward the posterior distribution of each fixed-effect parameter. Importantly, quadratic age effects can occur due to pre-peak age-dependent improvements and/or post-peak age-dependent declines (senescence). A priori planned post hoc analyses were performed for any reproductive trait exhibiting quadratic effects to estimate pre- and post-peak age effects (Bouwhuis et al., 2009; Keller, Reid, & Arcese, 2008; Reid, Bignal, Bignal, McCracken, & Monaghan, 2003). This was achieved by replacing the quadratic effect of age from the base model for two new fixed effects: (a) a binary variable 'pre-peak' (coded '0' for post-peak ages and '1' for pre-peak ages) and (b) the interaction between linear age and pre-peak. The main effect of age in this post hoc model represents the post-peak age effect while the interaction estimates the pre-peak age effect as a deviation from the post-peak age effect; the sum of the two represents the pre-peak age effect.

Models fitting parabolic age effects enable the calculation of reproductive peaks, but also force symmetrical pre- versus post-peak effects. If pre- and post-peak effects are not symmetrical, estimates of reproductive peaks may become biased. Fortunately, for the two traits showing nonlinear age effects (clutch size and fledging number in non-failed broods), pre- versus post-peak effects of age (which our post hoc model, detailed above, estimated independently) were relatively symmetrical (see Results and Table 1). Moreover, a version of Table 1 including the third-order effect of age showed that this effect was supported neither for clutch size (mean ± 95% credible interval (CI): 0.00, −0.03 to 0.02) nor for fledging number (0.03, −0.02 to 0.09). This implies that parabolic models seemed appropriate. We further tested whether the single age category with <5 data points (age = 7; see Results) biased our estimates (see Nussey, Kruuk, Donald, Fowlie, & Clutton-Brock, 2006 for a similar approach). Wethus re-ran our main analyses (Table 1) after combining ages 6 and 7, which did not change our estimates (Appendix S5 and Table S5).

Expanded models: estimating effects of individual-level exploratory behaviour

We expanded our base models to test whether within-individual age effects on reproduction varied with an individual's average value for exploratory behaviour (defined below). We did so by first estimating sources of variation in exploratory behaviour by fitting a univariate mixed-effects model with a fixed and random effects structure as detailed above (Table 1),

after which we simulated (using the r-package arm, see below) each individual's best linear unbiased predictor (BLUP) 1,000 times, and defined an individual's average value for exploratory behaviour as its mean BLUP over all simulations (i.e. producing one BLUP per individual). From previous work, we know that great tits habituate when repeatedly subjected to the novel environment test; in this and other (Dingemanse et al., 2012), great tit datasets, age and inter-year test sequence are fully conflated by design (i.e. surviving birds are subjected to repeated tests when older). Pragmatically fitting age (though functionally hard to interpret) thus enabled us to avoid bias in our estimates of individual-specific average values. Next, we expanded our base models by including each individual's average level (BLUP) of exploratory behaviour as a mean and variance-standardized covariate; we then fitted its interaction with each of the four age variables (i.e. age, age squared, first and last observed age) (Table 2). The usage of BLUPs as covariates has been criticized when uncertainty associated with BLUPs is not taken forward (Hadfield, Wilson, Garant, Sheldon, & Kruuk, 2010; Houslay & Wilson, 2017). Appendix S6 describes simulations demonstrating that taking forward uncertainty in BLUP values resulted in biased estimates; fitting average BLUP values instead produced estimates that were less precise yet unbiased; average BLUP values were therefore used throughout.

Model implementation

Statistical analyses were carried out using the packages ‘lme4’ and ‘arm’ in R-v3.3.2 (R Development Core Team, 2017). Model fit was assessed by visual inspection of the residuals (see Appendix S1). Based on 5,000 simulations, we extracted the 95% CIs (Gelman & Hill, 2007), representing the uncertainty around our estimates. Assessment of statistical support was thus obtained from the posterior distribution of each parameter, simulated using the sim function. We considered an effect ‘strongly supported’ if zero was not included within the 95% CI, and ‘moderately supported’ if the point estimate was skewed away from zero while its 95% CI simultaneously overlapped zero. Estimates centred on zero were viewed as strong support for the absence of an effect.

RESULTS

We acquired reproductive data for 599 (age = 1; 49.5% of all broods), 379 (age = 2; 31.3%), 151 (age = 3; 12.5%), 58 (age = 4; 4.8%), 16 (age = 5; 1.3%), 5 (age = 6; 0.4%) and 1 (age = 7; 0.1%) annual first clutches. For 95% (1,154 of 1,209), we assayed female exploratory behaviour, which we subsequently used to calculate a single average value (see Methods) for

each individual over all its assays; average exploratory behaviour was therefore available for 98% (1,187) of all clutches.

Exploratory behaviour

Exploratory behaviour dropped from 70.11 hops per (2-min) assay in 1-year-olds (intercept value; Table 1) with 3.89 hops per assay per year of age (negative effect of linear age; Table 1; Figure 1a); nonlinear age effects were not supported (quadratic age effect; Table 1). A first observed age effect was strongly supported (Table 1). Specifically, females first breeding when 2 years old or older (age ≥ 2) were behaving faster than birds recruiting as 1-year-olds (Figure 1a). Females were moderately repeatable in behaviour across years: adjusted individual across year repeatability (r) was 0.32 (Table 1). Plot, year and plot-year identity explained little variation if any at all (Table 1).

Clutch size

Clutch size varied within the average female as a function of linear and quadratic age (Table 1; Figure 1b). Clutch size was highest for 3-year-olds (age at peak: 3.4; Table 1). Before the age of peak performance, clutch size increased with 0.20 eggs per year of age (pre-peak age effect). Afterwards, clutch size decreased with 0.24 eggs per year (post-peak age effect); this decrease was moderately supported (Table 1). Clutch size thus showed age-dependent improvements that diminished with age, likely followed by a post-peak decline due to reproductive senescence.

Female exploratory behaviour predicted how clutch size varied with age. The main effect of exploratory behaviour centred on zero (Table 2); because we left-centred age (see Methods), this implied that exploratory behaviour did not affect clutch size among 1-yearolds. Instead, exploratory behaviour affected subsequent changes with age: exploratory behaviour interacted with both linear (moderate support) and quadratic (strong support) age (Table 2). Plots of parameter estimates for linear (Figure 2a) and quadratic (Figure 2b) age effects as a function of exploratory behaviour visualized the statistical nature of these interactions. These plots implied that the slowest half of females (values < 0) did not change clutch size with age: their parameter estimates for linear (Figure 2a) and quadratic (Figure 2b) age centred on zero. Consequently, the 50% slowest explorers produced moderate-sized clutches throughout their reproductive lives (Figure 3a, raw data controlling for random effects; Figure 3c, model predictions). By contrast, there was strong support for the fastest half (values ≥ 0) to exhibit age-dependent clutch sizes: credible intervals for this group did not overlap zero for either linear (Figure 2a) or quadratic (Figure 2b) age effects. These 50% fastest explorers improved

Table 1. Sources of variation in clutch size, number and average mass of fledglings (for nests producing any fledglings), and exploratory behaviour

	Clutch Size Count	No. Fledglings Count	Fledgling Mass Grams	Exploratory activity Count (No. hops)
Fixed effects	**β (95 CI)**	**β (95 CI)**	**β (95 CI)**	**β (95 CI)**
Intercept[1]	8.14 (7.8, 8.48)	5.36 (4.66, 6.04)	15.13 (14.5, 15.76)	70.11 (66.6, 73.64)
Linear age	0.35 (0.17, 0.54)	0.41 (0.07, 0.77)	0.06 (-0.23, 0.37)	-3.89 (-7.05, -0.79)
Quadratic age	-0.08 (-0.13, -0.03)	-0.14 (-0.24, -0.05)	-0.01 (-0.08, 0.07)	-0.05 (-0.85, 0.76)
First age	-0.13 (-0.39, 0.14)	-0.01 (-0.34, 0.32)	0.2 (-0.1, 0.52)	5.53 (1.98, 9.16)
Last age	-0.02 (-0.14, 0.09)	0.04 (-0.1, 0.18)	0 (-0.13, 0.13)	-0.04 (-1.52, 1.52)
Random effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual	1.48 (1.36, 1.61)	0.22 (0.19, 0.26)	0.43 (0.37, 0.49)	143.75 (129.3, 159.83)
Plot x Year	0.04 (0.03, 0.05)	0.26 (0.19, 0.35)	0.33 (0.24, 0.43)	0.78 (0.55, 1.04)
Plot	0.07 (0.03, 0.12)	0.24 (0.09, 0.46)	0.21 (0.09, 0.39)	13.53 (5.2, 25.58)
Year	0.14 (0.08, 0.24)	0.77 (0.34, 1.46)	0.59 (0.27, 1.09)	11.29 (4.14, 23.38)
Residual	0.83 (0.76, 0.9)	3.02 (2.75, 3.31)	2.23 (2.04, 2.44)	275.99 (253.82, 298.84)
Adjusted Repeatability	***r* (95% CI)**	***r* (95% CI)**	***r* (95% CI)**	***r* (95% CI)**
Individual	0.58 (0.55, 0.61)	0.05 (0.04, 0.06)	0.11 (0.1, 0.13)	0.32 (0.3, 0.35)
Plot x Year	0.02 (0.01, 0.02)	0.06 (0.04, 0.08)	0.09 (0.06, 0.11)	0 (0, 0)
Plot	0.03 (0.01, 0.05)	0.05 (0.02, 0.1)	0.06 (0.02, 0.1)	0.03 (0.01, 0.06)
Year	0.06 (0.03, 0.09)	0.17 (0.08, 0.28)	0.15 (0.08, 0.26)	0.03 (0.01, 0.05)
Residual	0.32 (0.3, 0.35)	0.67 (0.58, 0.75)	0.59 (0.52, 0.65)	0.62 (0.59, 0.65)
Peak performance	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**
Trait value at peak[2]	8.56 (8.14, 8.98)	5.67 (4.9, 6.44)	NA	NA
Age at peak[3]	2.4 (1.66, 3.68)	1.4 (0.5, 2.16)	NA	NA
Pre/post-peak analysis	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**
Pre-peak age effect[4]	0.2 (0.08, 0.32)	0.36 (0.01, 0.72)	NA	NA
Post-peak age effect[5]	-0.24 (-0.6, 0.13)	-0.51 (-0.89, -0.12)	NA	NA
Sample sizes	**n**	**n**	**n**	**n**
PlotYear	84	84	84	84
Plot	12	12	12	12
Year	7	7	7	7
Individual	813	625	671	791
Observations	1209	894	962	1154

[1]First-time breeders (i.e., second-year birds) were given age equal to zero; for all traits except clutch size, brood size manipulation category (see Material and Methods) was fitted as an additional fixed-effect factor with "not manipulated" set as the reference category (see Supplementary Text S3); estimates for other treatment groups (1 = control; 2 = enlarged; 3 = reduced) are printed in Supplementary Table S3. The statistical intercept is therefore for second-year birds and nests that were not manipulated.

[2]Calculated as $\beta_0 - \beta_1^2/4\beta_2$, where β_0 = the statistical intercept, β_1 = age (linear term), β_2 = age (quadratic tem); not calculated for traits failing to exhibit significant quadratic age effects ("NA").

[3]Calculated as $-\beta_1/2\beta_2$, where β_1 = age (linear term), β_2 = age (quadratic term); not calculated for traits failing to exhibit significant quadratic age effects ("NA").

[4]Effect of linear age before peak performance (post-hoc analysis; detailed in the Methods)

[5]Effect of linear age after peak performance (post-hoc analysis; detailed in the Methods)

clutch size with age in a diminishing fashion, possibly followed by an age-dependent decline (i.e. reproductive senescence) when old (Figure 3b, raw data controlling for random effects;

Figure 3d, model predictions). We came to the same conclusion when we re-ran our models with the same random and fixed-effect structure as printed in Table 1 but separately for each of the two groups. In the slowest half, neither linear (parameter estimate with 95% CIs: –0.06, −0.35 to 0.23) nor quadratic (0.04, −0.05 to 0.12) effects of age were supported (Figure 3c); by contrast, in the fastest half of the females, linear (0.62, 0.37–0.85) and quadratic (–0.13, –0.18 to –0.07) effects of age were both supported (Figure 3d).

Importantly, the distribution of ages differed between the 50% slowest versus fastest explorers. The slowest explorers only had reproductive data for 1- to 5-year-olds (ages 1–5: n = 293, 187, 73, 28, 8 first clutches), the fastest explorers instead for 1- to 7-year-olds (ages 1–7: n = 295, 182, 77, 30, 8, 5, 1). A follow-up analysis using only ages where both had data (i.e. excluding n = 6 data points of age > 5) resulted in the same level of support for interactive effects between exploratory behaviour and linear and quadratic age (Table S7). Thus, our finding of personality-related age dependency of clutch size was not an artefact caused by lack of data for older slow explorers. Note that those post hoc analyses of discrete groups (slower vs. faster explorers) enabled us to interpret, and verbally present, complex interaction terms between continuous predictors (age and exploratory behaviour), and should not be taken as evidence for the existence of two discrete forms of age-dependent clutch sizes within the population.

Annual fledgling number and average mass

Annual fledgling number (in non-failing broods) varied within individuals with both linear and quadratic age (Table 1; Figure 1b). Peak performance occurred when birds were between 2 and 3 years old (Table 1). Before the age of peak performance, fledgling number increased with 0.36 offspring per year of age (pre-peak age effects; Table 1). Afterwards, it decreased with 0.51 offspring per year of age (post-peak age effects; Table 1). Annual fledging number showed age-dependent improvements with increased breeding experience (pre-peak age effect), followed by an age-dependent decline due to reproductive senescence (post-peak age effect) that was strongly supported.

Point estimates for interactions between linear (or quadratic) age and exploratory behaviour suggested that the same pattern of personality-related age-dependent reproduction described above for clutch size also characterized fledging number (Table 2). For fledgling number, however, the support was moderate at best owing to skewed 95% CIs (particularly for exploratory behaviour × quadratic age) that nevertheless included zero.

Average fledging mass did not vary with linear or quadratic age (Table 1; Figure 1d), neither did those effects vary as a function of female exploratory behaviour (Table 2).

Selective (dis)appearance

We detected no evidence for selective (dis)appearance effects: first and last observed age of reproduction effects were not supported (Table 1). Our expanded analyses showed that first observed age effects were not supported for birds of average exploratory behaviour (main effect of first observed age; Table 2), echoing results of our main analyses (Table 1). However, there was strong support for a first observed age effect to decrease with increasing exploratory behaviour (interaction first observed age × exploratory behaviour; Table 2). Inspection of the raw data suggested this interaction resulted from faster—but not slower—explorers exhibiting decreased clutch sizes when they were older than first-year-olds at first observed breeding (Figure 3a, b).

Table 2. Effects of individual exploratory behaviour on within-individual age dependency of reproductive traits: clutch size, and number and average mass of fledglings (for nests producing any fledglings)

	Clutch Size Count	No. Fledglings Count	Fledgling Mass Grams
Fixed effects	**β (95 CI)**	**β (95 CI)**	**β (95 CI)**
Intercept	8.15 (7.82, 8.49)	5.34 (4.62, 6.05)	15.14 (14.52, 15.75)
Linear age	0.31 (0.12, 0.49)	0.37 (-0.01, 0.74)	0.1 (-0.21, 0.42)
Quadratic age	-0.06 (-0.11, -0.01)	-0.12 (-0.22, -0.02)	-0.02 (-0.1, 0.06)
First age	-0.11 (-0.38, 0.17)	0 (-0.36, 0.35)	0.17 (-0.13, 0.48)
Last age	-0.03 (-0.14, 0.09)	0.04 (-0.11, 0.18)	-0.01 (-0.13, 0.12)
Exploration	-0.04 (-0.2, 0.12)	0.08 (-0.13, 0.31)	0.01 (-0.18, 0.2)
Exploration x Linear age	0.15 (0, 0.3)	0.05 (-0.24, 0.37)	-0.08 (-0.35, 0.18)
Exploration x Quadratic age	-0.05 (-0.09, -0.01)	-0.04 (-0.11, 0.04)	0.02 (-0.05, 0.09)
Exploration x First age	-0.41 (-0.69, -0.14)	0 (-0.33, 0.32)	0.26 (-0.03, 0.57)
Exploration x Last age	0.08 (-0.01, 0.18)	0.01 (-0.11, 0.13)	0.02 (-0.09, 0.13)
Random effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual	1.49 (1.37, 1.62)	0.23 (0.2, 0.26)	0.4 (0.35, 0.46)
Plot x Year	0.04 (0.03, 0.05)	0.26 (0.19, 0.35)	0.34 (0.25, 0.45)
Plot	0.07 (0.03, 0.12)	0.26 (0.1, 0.51)	0.19 (0.08, 0.37)
Year	0.14 (0.08, 0.24)	0.77 (0.35, 1.58)	0.53 (0.27, 1.03)
Residual	0.82 (0.76, 0.89)	3.01 (2.74, 3.32)	2.27 (2.07, 2.49)
Adjusted Repeatability	***r* (95% CI)**	***r* (95% CI)**	***r* (95% CI)**
Individual	0.58 (0.55, 0.61)	0.05 (0.04, 0.06)	0.11 (0.09, 0.12)
Plot x Year	0.02 (0.01, 0.02)	0.06 (0.04, 0.08)	0.09 (0.07, 0.12)
Plot	0.03 (0.01, 0.05)	0.06 (0.02, 0.11)	0.05 (0.02, 0.09)
Year	0.05 (0.03, 0.09)	0.17 (0.09, 0.3)	0.14 (0.08, 0.24)
Residual	0.32 (0.3, 0.34)	0.66 (0.56, 0.74)	0.61 (0.53, 0.66)

Note: Fixed and random parameters are detailed in Table 1. We print here our expanded models that include an individual's estimated average exploratory behaviour ('Exploration', representing the individual's best linear unbiased predictor derived from the analysis printed in Table 1), and its interactions with all age variables, as additional fixed effects.

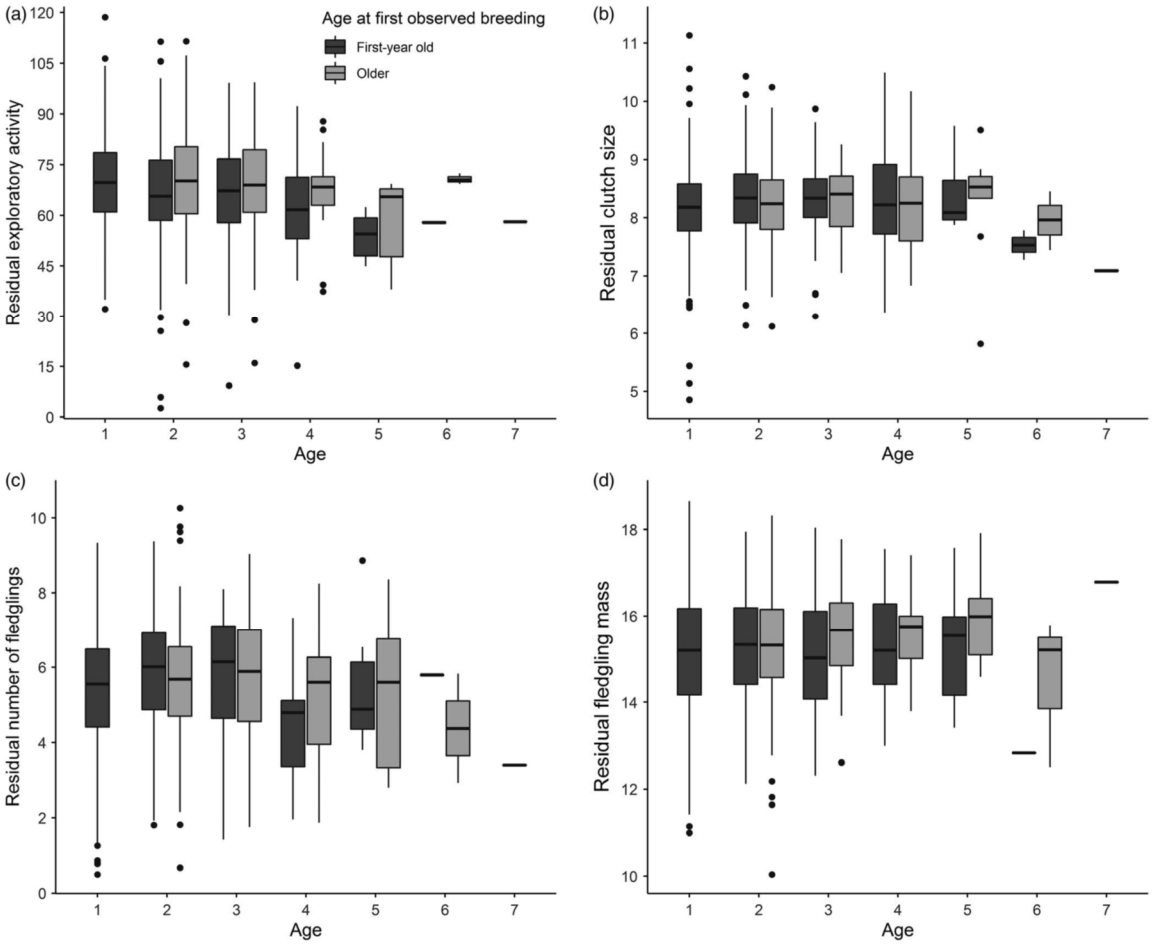


Figure 1. Box plots per age class for (a) exploratory behaviour, (b) clutch size, (c) number of fledglings (for non failed nests) and (d) average fledgling mass. Plotted are residuals from a model controlling solely for random effects listed in Table 1. Separate box plots for birds with first observed age equal to one year old versus older.

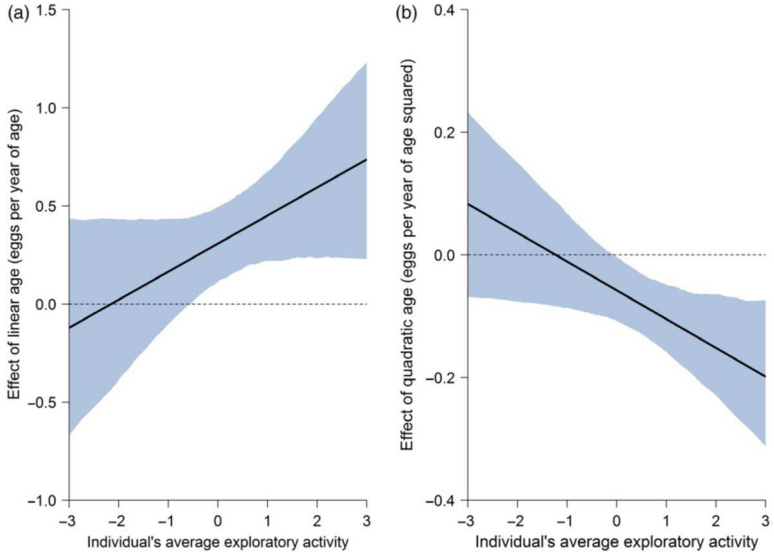


Figure 2. The within-individual effect of (a) linear and (b) quadratic age on clutch size (eggs per year of age) as a function of an individual's average exploratory behaviour. The black line represents the point estimate with 95% credible intervals (CIs; blue shaded area) derived from the analysis printed in Table 2. Linear and quadratic age effects were supported only for the 50% fastest explorers (values ≥ 0) and were, respectively, positive versus negative.

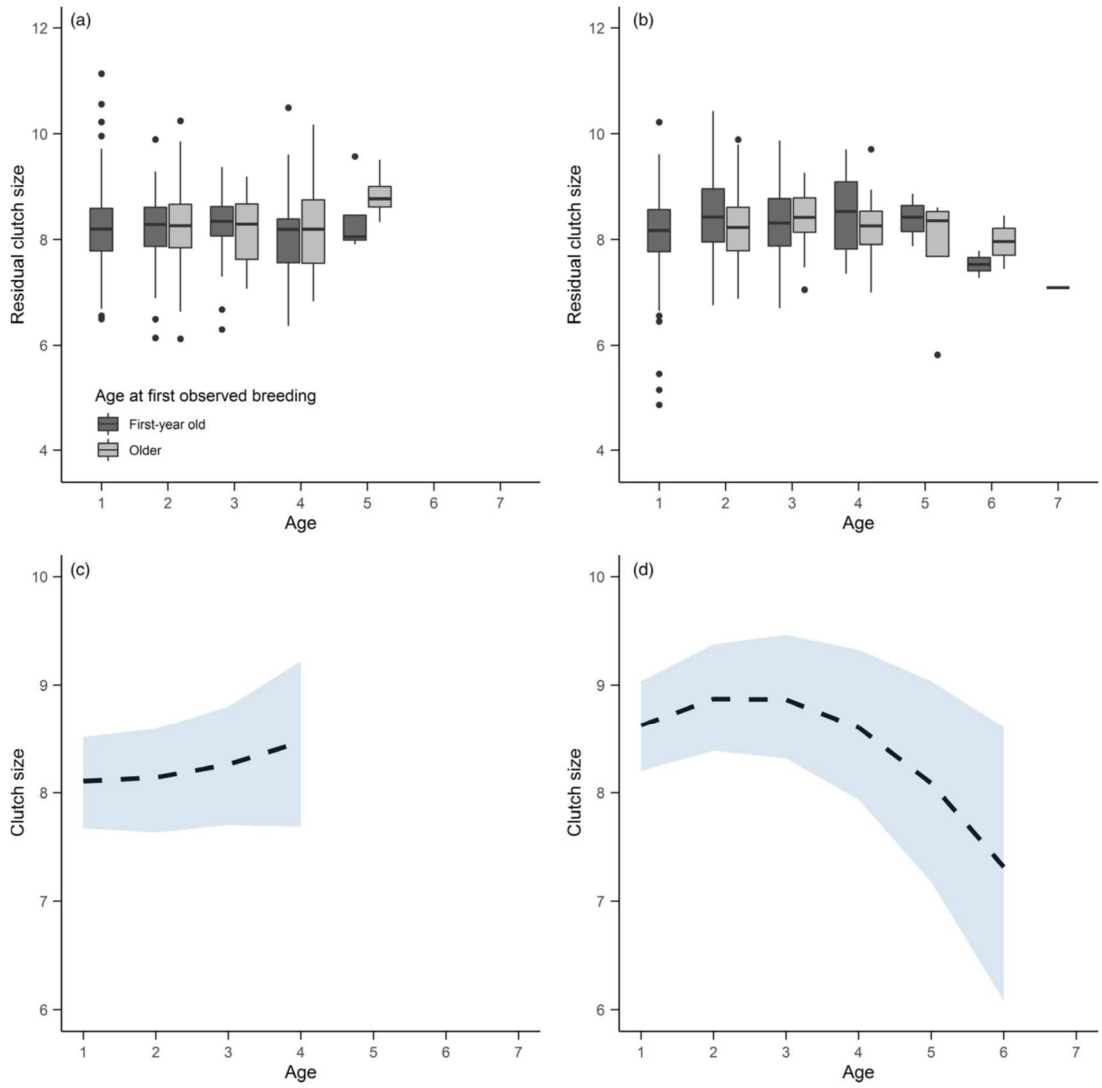


Figure 3. Personality-related age dependency of clutch size. We show box plots per age class for the 50% (a) slowest versus (b) fastest explorers; we plot residuals from a model controlling for random effects listed in Table 1, with separate box plots for first observed age equal to one year old versus older. We also plotted the average pattern of within-individual age dependency of clutch size within the (c) 50% slowest versus (d) fastest explorers; the black line represents the point estimate with 95% credible intervals (CIs; blue shaded area) derived from the analysis printed in Table 2

DISCUSSION

Optimal behavioural phenotypes should vary with how life-history trade-offs are resolved (Réale et al., 2010; Ricklefs & Wikelski, 2002; Wolf et al., 2007). Adaptive theory predicts that aggressive, bold or explorative individuals trade off future for current reproduction, leading to a faster pace-of-life (Dammhahn et al., 2018; Mathot & Frankenhuis, 2018; Réale et al., 2010). Previous tests utilizing life span as a proxy for allocation to future reproduction failed to overall support pace-of-life syndrome (POLS) theory (meta-analyses: Royauté et al., 2018; Tarka, Guenther, Niemelä, Nakagawa, & Noble, 2018). We identified here within-individual patterns of age-dependent reproduction, and potentially reproductive senescence, as

key components of life history varying with individual risky behaviour. Specifically, slower explorers produced moderate-sized clutches throughout their reproductive lives, showing neither evidence for age-related improvements when young nor evidence for age-related declines when older (Figure 3a, c), though we note that data for old age classes were not available for slower explorers. By contrast, over the same range of age classes as observed for slower explorers (1- to 4-year-olds), faster explorers instead showed age-related improvements that diminished with age (Figure 3b, d). There was moderate support for faster explorers subsequently showing reproductive senescence, though this evidence should be taken with caution as it is based on little data. Importantly, the same pattern may have characterized annual fledgling success, if so, unpredictable environmental effects diluted personality-related differences in this downstream reproductive trait (see also Hutfluss & Dingemanse, 2019 for a similar finding and further discussion). Overall, future studies should consider reproductive senescence as a key component of life history mediating personality-related differences in how trade-offs between current and future reproduction are resolved.

First-year-olds produced moderate-sized clutches regardless of exploration type. Faster explorers subsequently showed age-related increases in clutch size that lasted until they were 3-year-olds (Figure 3b). The majority of breeding records (93.3%) were for birds breeding as 1-year (49.5%), 2-year (31.3%) or 3-year-olds (12.5%), implying that faster explorers produced, on average, larger clutches than slower explorers for most of their reproductive lives; very few faster explorers thus lived long enough to experience reproductive declines at old age. Importantly, faster explorers cannot be shown to not have a shorter life span in this (Wischhoff & Dingemanse, In Preparation) or other great tit populations (Nicolaus et al., 2016). Slower explorers thus differed from faster ones in two important ways. First, only faster explorers showed (nonlinear) age-related increases in clutch size, likely followed by reproductive senescence. Second, faster explorers produced larger clutches for most of their reproductive life compared to slower explorers. If these age-related increases in clutch size observed in faster explorers represented an investment trading off with future reproduction, an assumption warranting experimental confirmation (Nicolaus et al., 2015), the moderately supported evidence for reproductive senescence among faster explorers may imply that they paid the costs of reproduction by reducing investment in physiological and cellular functioning in late life (see Introduction). Importantly, a recent simulation study implied that POLS-related variation in life-history traits measured once (e.g. longevity) will be extremely difficult to demonstrate empirically compared to POLS-related variation in life-history traits expressed repeatedly (e.g. clutch size) (Araya-Ajoy et al., 2018). We therefore need to be somewhat cautious in

interpreting publications failing to recover patterns of reduced longevity among faster explorers from empirical data.

Selective (dis)appearance and variation in experience

In this paper, we estimated within-individual patterns of age-dependent reproduction while controlling for potential biases resulting from within-individual behavioural plasticity and selective (dis)appearance of high- versus low-quality individuals. Females were moderately repeatable in reproductive traits; individuals of superior 'quality' (defined statistically as females with high intercepts for reproductive traits) might thus, for example, have recruited into the breeding population younger ('selective appearance'; first observed age effect), and/ or disappeared when older ('selective disappearance'; last observed age effect) (Bouwhuis et al., 2009). For a conceptual illustration of the idea, see Figure 1 in van de Pol and Verhulst (2006).

A first observed age effect explained variation in exploratory behaviour (Table 1; Figure 1a). Specifically, females first breeding as 1-year-olds were slower than females first breeding at older ages (Figure 1a). We offer two potential explanations. First, faster (vs. slower) explorers might recruit into the breeding population at an older age (selective appearance); this might explain why 'late' recruits produced, on average, 5.53 more hops (Table 1). Alternatively, late recruits might have hopped more because they lacked at least 1 year of experience with the assay. This latter explanation seemed more fitting because exploratory behaviour decreased with 3.89 hops per year of age (=experience; see Methods) within individual females, implying that 'late' recruits (lacking 1 year of experience) should hop more. Indeed, the effect of first observed age was not supported when this differential experience was statistically accounted for (Appendix S8).

Along the same lines, our analyses strongly supported an interactive effect of first observed age and exploratory behaviour on clutch size. This pattern did not imply personality-related selective appearance in the breeding population. Briefly, we observed age-related increases in clutch size solely for faster explorers (Figure 3b, d). A negative interaction between first observed age and exploratory behaviour on clutch size should thus emerge if such effects were attributable to breeding experience rather than age per se: 'late' recruiting faster explorers should lack breeding experience and thus produce smaller clutches. The interactive effect of first observed age and exploratory behaviour on clutch size thus does not constitute sound evidence for personality-related selective appearance; rather, it was expected because breeding experience (i.e. plasticity) affects reproductive performance.

In summary, while we did not find convincing evidence for selective (dis)appearance, we did learn that controlling for first observed age of reproduction provided a means to statistically control for individual differences in age-related experience. For example, it enabled us to conclude that the smaller clutch sizes produced by faster explorers recruiting at an older age were expected based on increases in clutch size with breeding experience. Moreover, exploratory behaviour varied with age and/or experience within individuals, implying that our concerns regarding effects of within-individual plasticity biasing estimates of personality-related age-dependent reproduction (see Introduction) were valid. Future studies should thus carefully consider multiple alternative explanations when interpreting age-related patterns in reproduction.

CONCLUSIONS

We demonstrated for a natural bird population that slower and faster explorers produced moderate-sized clutch sizes when young, after which faster explorers increased nonlinearly, peaked and likely decreased their clutch sizes while ageing, while slower explorers produced moderate-sized clutches throughout. Age-related reproduction thus represents a key component of POLSs. Certain parameters, particularly estimates of the age of peak reproduction or post-peak declines in reproductive performance, were, notably, based on relatively few data, particularly among older age classes. Those estimates are therefore relatively uncertain and warrant validation with larger samples. Experimental studies are further required to reveal whether trade-offs indeed underpin the covariance between life-history traits and risky personality identified in this paper.

ACKNOWLEDGEMENTS

N.J.D. was supported by the Max Planck Society (2012–2017) and the German Science Foundation (grant no. DI 1694/1-1). M.M. was funded by a Marie Skłodowska-Curie Individual Fellowship (PLASTIC TERN, Grant Agreement Number 793550). We thank past members of the Evolutionary Ecology of Variation group (Max Planck Institute for Ornithology) and past and current members of the Behavioural Ecology (LMU) group, Alexander Hutfluss, and all field assistants and students that helped collecting the data. We also thank members of the Evolutionary Ecology of Variation Group at North Dakota University, Pat Monaghan, Samantha Patrick, Neil Metcalfe, John Quinn and two anonymous reviewers for input. All procedures complied with guidelines from the District Government of Upper Bavaria (Regierung von Oberbayern) for Animal Care.

REFERENCES

Abbey-Lee, R. N., & Dingemanse, N. J. (2019). Adaptive individual variation in phenological responses to perceived predation levels. Nature Communications, 10, 1601. https://doi.org/10.1038/s41467-019-09138-5

Abbey-Lee, R., Mathot, K. J., & Dingemanse, N. J. (2016). Behavioral and morphological responses to perceived predation risk: A field experiment in passerines. Behavioral Ecology, 27, 857–864. https://doi.org/10.1093/behec o/arv228

Adriaenssens, B., Pauliny, A., Blomqvist, D., & Johnsson, J. I. (2016). Telomere length covaries with personality in wild brown trout. Physiology and Behavior, 165, 217–222. https://doi.org/10.1016/j.physb eh.2016.07.005

Araya-Ajoy, Y. G., Bolstad, G. H., Brommer, J., Careau, V., Dingemanse, N. J., & Wright, J. (2018) Demographic measures of an individual's "pace of life": fecundity rate, lifespan, generation time, or a composite variable? Behavioral Ecology and Sociobiology, 72(5), 75. https://doi.org/10.1007/s00265-018-2477-7

Araya-Ajoy, Y. G., & Dingemanse, N. J. (2017). Repeatability, heritability, and age dependence of seasonal plasticity in aggressiveness in a wild passerine bird. Journal of Animal Ecology, 86, 227–238. https://doi.org/10.1111/1365-2656.12621

Araya-Ajoy, Y. G., Kuhn, S., Mathot, K. J., Mouchet, A., Mutzel, A., Nicolaus, M., Dingemanse, N. J. (2016). Sources of (co)variation in alternative siring routes available to male great tits (Parus major). Evolution, 70, 2308–2321.

Bateson, M., Brilot, B. O., Gillespie, R., Monaghan, P., & Nettle, D. (2015). Developmental telomere attrition predicts impulsive decision-making in adult starlings. Proceedings of the Royal Society B: Biological Sciences, 282, 20142140. https://doi.org/10.1098/rspb.2014.2140

Bell, A. M., Hankison, S. J., & Laskowski, K. L. (2009). The repeatability of behaviour: A meta-analysis. Animal Behaviour, 77, 771–783. https://doi.org/10.1016/j.anbehav.2008.12.022

Biro, P. A., & Stamps, J. A. (2008). Are animal personality traits linked to life-history productivity? Trends in Ecology and Evolution, 23, 361– 368. https://doi.org/10.1016/j.tree.2008.04.003

Biro, P. A., & Stamps, J. A. (2010). Do consistent individual differences in metabolic rate promote consistent individual differences in behavior? Trends in Ecology and Evolution, 25, 653–659. https://doi.org/10.1016/j.tree.2010.08.003

Bouwhuis, S., Sheldon, B. C., Verhulst, S., & Charmantier, A. (2009). Great tits growing old: Selective disappearance and the partitioning of senescence to stages within the breeding cycle. Proceedings of the Royal Society B-Biological Sciences, 276, 2769–2777. https://doi.org/10.1098/rspb.2009.0457

Brommer, J. E., & Class, B. (2015). The importance of genotype-by-age interactions for the development of repeatable behavior and correlated behaviors over lifetime. Frontiers in Zoology, 12, S2. https://doi.org/10.1186/1742-9994-12-S1-S2

Brommer, J., Rattiste, K., & Wilson, A. (2010). The rate of ageing in a long-lived bird is not heritable. Heredity, 104, 363–370. https://doi.org/10.1038/hdy.2009.125

Brommer, J. E., Wilson, A. J., & Gustafsson, L. (2007). Exploring the genetics of aging in a wild passerine bird. The American Naturalist, 170, 643–650. https://doi.org/10.1086/521241

Bronshtein, I. N., Semendyayev, K. A., Musiol, G., & Mühlig, H. (2015). Handbook of mathematics. Berlin, Heidelberg: Springer-Verlag.

Careau, V., Thomas, D., Humphries, M. M., & Réale, D. (2008). Energy metabolism and animal personality. Oikos, 117, 641–653. https://doi.org/10.1111/j.0030-1299.2008.16513.x

Carter, A. J., Feeney, W. E., Marshall, H. H., Cowlishaw, G., & Heinsohn, R. (2013). Animal personality: What are behavioural ecologists measuring? Biological Reviews, 88, 465–475. https://doi.org/10.1111/brv.12007

Class, B., & Brommer, J. E. (2016). Senescence of personality in a wild bird. Behavioral Ecology and Sociobiology, 70, 733–744. https://doi.org/10.1007/s00265-016-2096-0

Dammhahn, M., Dingemanse, N. J., Niemelä, P. T., & Reale, D. (2018). Pace-of-life syndromes: a framework for the adaptive integration of personality and life-history. Behavioral Ecology and Sociobiology, 72, 62.

Dingemanse, N. J., Bouwman, K. M., van de Pol, M., van Overveld, T., Patrick, S. C., Matthysen, E., & Quinn, J. L. (2012). Variation in personality and behavioural plasticity across four populations of the great tit Parus major. Journal of Animal Ecology, 81, 116–126. https://doi.org/10.1111/j.1365-2656.2011.01877.x

Dingemanse, N. J., Moirón, M., Araya-Ajoy, Y. G., Mouchet, A., & Abbey- Lee, R. N. (2019). Data from: Individual variation in age-dependent reproduction: Fast explorers live fast but senesce young? Dryad Digital Repository, https://doi.org/10.5061/dryad.12jm6 3xss

Dingemanse, N. J., & Wolf, M. (2010). A review of recent models for adaptive personality differences. Philosophical Transactions of the Royal Society B-Biological Sciences, 365, 3947–3958.

Drent, P. J. (1984). Mortality and dispersal in summer and its consequences for the density of great tits Parus major at the onset of autumn. Ardea, 72, 127–162.

Evans, S. R., Gustafsson, L., & Sheldon, B. C. (2011). Divergent patterns of age-dependence in ornamental and reproductive traits in the collared flycatcher. Evolution, 65, 1623–1636. https://doi.org/10.1111/j.1558-5646.2011.01253.x

Fisher, D. N., David, M., Tregenza, T., & Rodriguez-Munoz, R. (2015). Dynamics of among individual behavioral variation over adult lifespan in a wild insect. Behavioral Ecology, 26, 975–985. https://doi.org/10.1093/behec o/arv048

Fisher, R. A. (1930). The genetical theory of natural selection. Oxford, UK: Oxford University Press.

Gelman, A., & Hill, J. (2007). Data analysis using regression and multi-variate/ hierarchical models. New York, NY: Cambridge University Press.

Hadfield, J. D., Wilson, A. J., Garant, D., Sheldon, B. C., & Kruuk, L. E. B. (2010). The misuse of BLUP in ecology and evolution. The American Naturalist, 175, 116–125. https://doi.org/10.1086/648604

Hamilton, W. D. (1966). Moulding of senescence by natural selection. Journal of Theoretical Biology, 12, 12–000.

Harvey, P. H., Greenwood, P. J., & Perrins, C. M. (1979). Breeding area fidelity of great tits (Parus major). Journal of Animal Ecology, 48, 305–313. https://doi.org/10.2307/4115

Holtmann, B., Lagisz, M., & Nakagawa, S. (2017). Metabolic rates, and not hormone levels, are a likely mediator of between-individual differences in behaviour: A meta-analysis. Functional Ecology, 31, 685–696. https://doi.org/10.1111/1365-2435.12779

Houle, D., Pelabon, C., Wagner, G. P., & Hansen, T. F. (2011). Measurement and meaning in biology. Quarterly Review of Biology, 86, 3–34. https://doi.org/10.1086/658408

Houslay, T. M., & Wilson, A. J. (2017). Avoiding the misuse in BLUP in behavioral ecology. Behavioral Ecology, 28, 948–952.

Hutfluss, A., & Dingemanse, N. J. (2019). Human recreation reduces clutch size in great tits Parus major regardless of risk-taking personality. Behavioral Ecology, 30, 1751–1760.

Jenni, L., & Winkler, R. (1994). Moult and ageing of European passerines. London, UK: Academic Press.

Jones, O. R., Gaillard, J. M., Tuljapurkar, S., Alho, J. S., Armitage, K. B., Becker, P. H., Coulson, T. (2008). Senescence rates are determined by ranking on the fast-slow life-history continuum. Ecology Letters, 11, 664–673. https://doi.org/10.1111/j.1461-0248.2008.01187.x

Keller, L. F., Reid, J. M., & Arcese, P. (2008). Testing evolutionary models of senescence in a natural population: Age and inbreeding effects on fitness components in song sparrows. Proceedings of the Royal Society B-Biological Sciences, 275, 597–604. https://doi.org/10.1098/rspb.2007.0961

Kluen, E., & Brommer, J. E. (2013). Context-specific repeatability of personality traits in a wild bird: A reaction-norm perspective. Behavioral Ecology, 24, 650–658. https://doi.org/10.1093/behec o/ars221

Lemaitre, J. F., Berger, V., Bonenfant, C., Douhard, M., Gamelon, M., Plard, F., & Gaillard, J. M. (2015). Early-late life trade-offs and the evolution of ageing in the wild. Proceedings of the Royal Society B: Biological Sciences, 282, 20150209. https://doi.org/10.1098/rspb.2015.0209

Mathot, K. J., & Frankenhuis, W. E. (2018). Models of pace-of-life syndromes (POLS): A systematic review. Behavioral Ecology and Sociobiology, 72, 41. https://doi.org/10.1007/s00265-018-2459-9

Mathot, K. J., Nicolaus, M., Araya-Ajoy, Y. G., Dingemanse, N. J., & Kempenaers, B. (2015). Does metabolic rate predict risk-taking behaviour? A field experiment in a wild passerine bird. Functional Ecology, 29, 239–249. https://doi.org/10.1111/1365-2435.12318

Medawar, P. B. (1952). An unsolved problem in biology. London, UK: H.K. Lewis & Co., Ltd..

Moiron, M., Araya-Ajoy, Y. G., Mathot, K. J., Mouchet, A., & Dingemanse, N. J. (2019). Functional relations between body mass and risk-taking behavior in wild great tits. Behavioral Ecology, 30, 617–623. https://doi.org/10.1093/behec o/ary199

Montiglio, P. O., Dammhahn, M., Messier, G. D., & Reale, D. (2018). The pace-of-life syndrome revisited: The role of ecological conditions and natural history on the slow-fast continuum. Behavioral Ecology and Sociobiology, 72, 9. https://doi.org/10.1007/s00265-018-2526-2

Nicolaus, M., Mathot, K. J., Araya-Ajoy, Y. G., Mutzel, A., Wijmenga, J. J., Kempenaers, B., & Dingemanse, N. J. (2015). Does coping style predict optimization: An experimental test in a wild passerine bird. Proceedings of the Royal Society B-Biological Sciences, 282, 20142405. https://doi.org/10.1098/rspb.2014.2405

Nicolaus, M., Piault, R., Ubels, R., Tinbergen, J. M., & Dingemanse, N. J. (2016). The correlation between colouration and exploration behaviour varies across hierarchical levels in a wild passerine bird. Journal of Evolutionary Biology, 29, 1780–1792.

Nicolaus, M., Tinbergen, J. M., Bouwman, K. M., Michler, S. P. M., Ubels, R., Both, C., Dingemanse, N. J. (2012). Experimental evidence for adaptive personalities in a wild passerine bird. Proceedings of the Royal Society B-Biological Sciences, 279, 4885–4892. https://doi.org/10.1098/rspb.2012.1936

Niemelä, P. T., & Dingemanse, N. J. (2018a). Meta-analysis reveals weak associations between intrinstic state and personality. Proceedings of the Royal Society B-Biological Sciences, 1873, 20172823.

Niemelä, P. T., & Dingemanse, N. J. (2018b). On the usage of single measurements in behavioural ecology research on individual differences. Animal Behavior, 145, 99–105. https://doi.org/10.1016/j.anbeh av.2018.09.012

Niemelä, P., Dingemanse, N. J., Alioravainen, N., Vainikka, A., & Kortet, R. (2013). Personality pace-of-life hypothesis: Testing genetic associations among personality and life-history. Behavioral Ecology, 24, 935–941. https://doi.org/10.1093/beheco/art014

Nussey, D. H., Kruuk, L. E. B., Donald, A., Fowlie, M., & Clutton-Brock, T. H. (2006). The rate of senescence in maternal performance increases with early-life fecundity in red deer. Ecology Letters, 9, 1342–1350. https://doi.org/10.1111/j.1461-0248.2006.00989.x

Patrick, S. C., Charmantier, A., & Weimerskirch, H. (2013). Differences in boldness are repeatable and heritable in a long-lived marine predator. Ecology and Evolution, 3, 4291–4299. https://doi.org/10.1002/ece3.748

Patrick, S. C., & Weimerskirch, H. (2015). Senescence rates and late adulthood reproductive success are strongly influenced by personality in a long-lived seabird. Proceedings of the Royal Society B-Biological Sciences, 282, 20141649. https://doi.org/10.1098/rspb.2014.1649

Pauliny, A., Devlin, R. H., Johnsson, J. I., & Blomqvist, D. (2015). Rapid growth accelerates telomere attrition in a transgenic fish. BMC Evolutionary Biology, 15, 159. https://doi.org/10.1186/s12862-015-0436-8

Réale, D., Garant, D., Humphries, M. M., Bergeron, P., Careau, V., & Montiglio, P. O. (2010). Personality and the emergence of the pace-of-life syndrome concept at the population level. Philosophical Transactions of the Royal Society of London Series B, 365, 4051–4063. https://doi.org/10.1098/rstb.2010.0208

Réale, D., Martin, J., Coltman, D. W., Poissant, J., & Festa-Bianchet, M. (2009). Male personality, life-history strategies and reproductive success in a promiscuous mammal. Journal of Evolutionary Biology, 22, 1599–1607. https://doi.org/10.1111/j.1420-9101.2009.01781.x

Reid, J. M., Bignal, E. M., Bignal, S., McCracken, D. I., & Monaghan, P. (2003). Age-specific reproductive performance in red-billed choughs Pyrrhocorax pyrrhocorax: Patterns and processes in a natural population. Journal of Animal Ecology, 72, 765–776. https://doi.org/10.1046/j.1365-2656.2003.00750.x

Ricklefs, R. E., & Wikelski, M. (2002). The physiology/life-history nexus. Trends in Ecology and Evolution, 17, 462–468. https://doi.org/10.1016/S0169-5347(02)02578-8

Royauté, R., Berdal, M., Hickey, C., & Dochtermann, N. A. (2018). Paceless life? A meta-analysis of the "pace-of-life syndrome". Behavioral Ecology and Sociobiology, 72, 64.

Saether, B. E. (1988). Pattern of covariation between life-history traits of European birds. Nature, 331, 616–617. https://doi.org/10.1038/331616a0

Santostefano, F., Wilson, A. J., Niemelä, P. T., & Dingemanse, N. J. (2017). Behavioural mediators of genetic life-history trade-offs in field crickets: A test of the pace-of-life syndrome hypothesis in field crickets. Proceedings of the Royal Society B: Biological Sciences, 284, 1864.

Smith, B. R., & Blumstein, D. T. (2008). Fitness consequences of personality: A meta-analysis. Behavioral Ecology, 19, 448–455. https://doi.org/10.1093/behec o/arm144

Stamps, J. A. (2007). Growth-mortality tradeoffs and 'personality traits' in animals. Ecology Letters, 10, 355–363. https://doi.org/10.1111/j.1461-0248.2007.01034.x

Stearns, S. C. (1992). The evolution of life histories. New York, NY: Oxford University Press.

Stuber, E. F., Araya-Ajoy, Y. G., Mathot, K. J., Mutzel, A., Nicolaus, M., Wijmenga, J. J., Dingemanse, N. J. (2013). Slow explorers take less risk: A problem of sampling bias in ecological studies. Behavioral Ecology, 24, 1092–1098. https://doi.org/10.1093/beheco/art035

Tarka, M., Guenther, A., Niemelä, P. T., Nakagawa, S., & Noble, D. W. A. (2018). Sex differences in life-history, behavior and physiology along a slow-fast continuum: A meta-analysis. Behavioral Ecology and Sociobiology, 72, 132. https://doi.org/10.1007/s00265-018-2534-2

van de Pol, M., & Verhulst, S. (2006). Age-dependent traits: A new statistical model to separate within- and between-individual effects. The American Naturalist, 167, 766–773. https://doi.org/10.1086/503331

van Noordwijk, A. J., McCleery, R. H., & Perrins, C. M. (1995). Selection for the timing of great tit breeding in relation to caterpillar growth and temperature. Journal of Animal Ecology, 64, 451–458. https://doi.org/10.2307/5648

Verbeek, M. E. M., Drent, P. J., & Wiepkema, P. R. (1994). Consistent individual differences in early exploratory behaviour of male great tits. Animal Behaviour, 48, 1113–1121. https://doi.org/10.1006/anbe.1994.1344

Williams, G. C. (1957). Pleiotropy, natural selection, and the evolution of senescence. Evolution, 11, 398–411. https://doi.org/10.1111/j.1558-5646.1957.tb02911.x

Williams, G. C. (1966). Natural selection costs of reproduction and a refinement of Lack's principle. The American Naturalist, 100, 687–690.

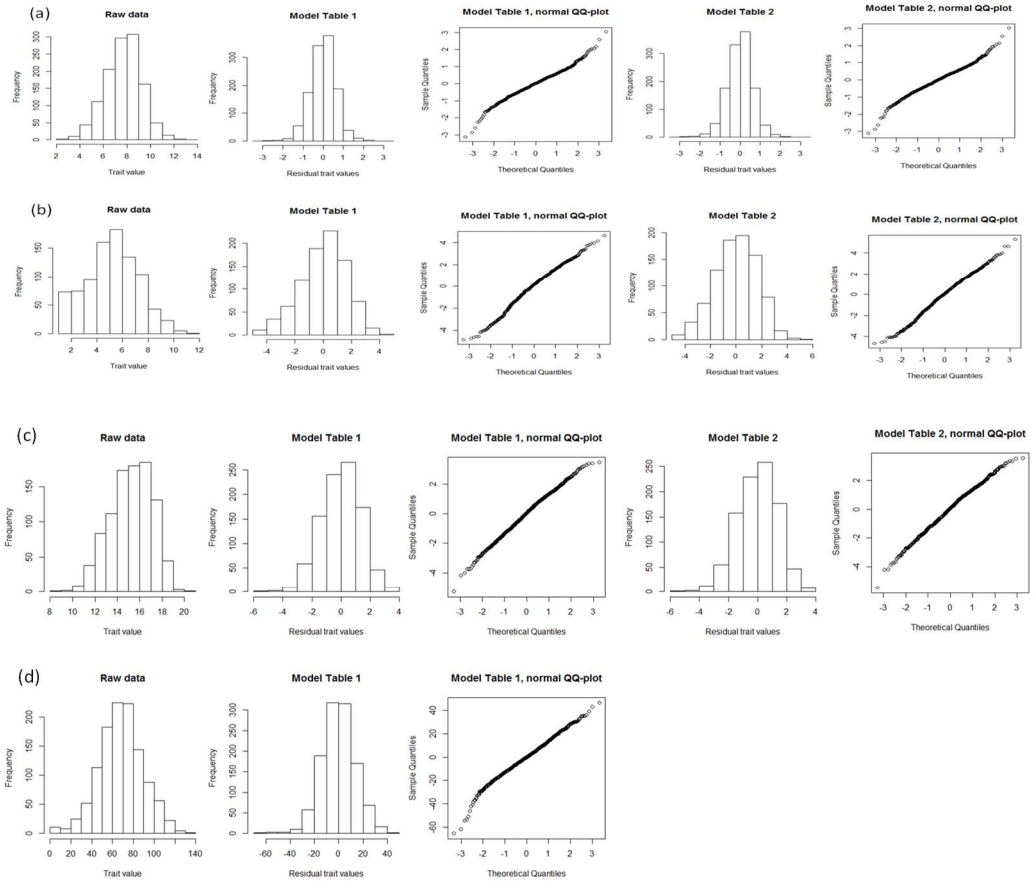
Wolf, M., van Doorn, G. S., Leimar, O., & Weissing, F. J. (2007). Life-history trade-offs favour the evolution of animal personalities. Nature, 447, 581–585. https://doi.org/10.1038/nature05835

SUPPORTING INFORMATION FOR CHAPTER 1

Appendix S1

We present here plots showing i) the distribution of the raw data, ii) the distribution of residuals derived from the model presented in Table 1, iii) a normal QQ-plot derived for the model presented in Table 1, and for reproductive traits only, iv) the distribution of residuals derived from the model presented in Table 2, and v) a normal QQ-plot derived for the model presented in Table 2 implied that distributional assumptions regarding normality were met for all traits analysed with Gaussian error distributions (Figure S1).

Figure S1. Plots showing i) the distribution of the raw data, ii) the distribution of residuals derived from the model presented in Table 1, iii) a normal QQ-plot derived for the model presented in Table 1, and for reproductive traits only, iv) the distribution of residuals derived from the model presented in Table 2, and v) a normal QQ-plot derived for the model presented in Table 2 for (a) clutch size, (b) number of fledglings in broods that did not fail, (c) offspring body mass in broods that did not fail, and (d) exploratory activity.



Appendix S2

The binary probability to produce any nestlings varied neither with linear nor with quadratic age within individual females (Table S2a). Last observed age of reproduction did, by contrast, explain variation: females with statistical intercept values indicative of a high likelihood to fledge any offspring disappeared from the population at a relatively old age (positive effect of last observed age; Table S1a). A female's average exploratory behaviour (BLUP) did not affect the binary probability to produce any fledglings (main effect), neither were interactions between exploratory activity and linear age, quadratic age, first observed age, or last observed age supported (Table S2b).

Table S2. Sources of variation in the probability to produce any fledglings. (**a**) Our base model. Fixed effect parameter estimates (β) with 95% credible intervals (CIs) are printed for linear (age) and quadratic (age × age) effects of absolute age (years) within individual females for a model controlling for effects of selective (dis)appearance by fitting first and last observed age. Individual, plot × year, plot, and year were fitted as random effects; variance attributable to each effect is printed both as an absolute value (σ^2) and as a proportion of the variance not attributable to random effects (adjusted repeatability, *r*). (**b**) Our expanded model, considering interaction effects between an individual's Best Linear Unbiased Predictor (BLUP) for exploratory behaviour ("exploration") and the age variables detailed above. Estimates were derived from a model fitting the bobyga optimizer.

	(a)	(b)
Fixed effects[1]	**β (95 CI)**	**β (95 CI)**
Intercept	0.89 (0.23, 1.53)	0.88 (0.24, 1.52)
Linear age	-0.18 (-0.64, 0.26)	-0.17 (-0.63, 0.29)
Quadratic age	-0.07 (-0.18, 0.04)	-0.08 (-0.19, 0.04)
First age	-0.18 (-0.5, 0.14)	-0.16 (-0.5, 0.16)
Last age	0.5 (0.29, 0.71)	0.52 (0.3, 0.73)
BSM: control (0)	0.15 (-0.69, 0.94)	0.05 (-0.79, 0.91)
BSM: enlarged (+3)	-0.4 (-1.14, 0.33)	-0.42 (-1.17, 0.34)
BSM: reduced (-3)	0.03 (-0.75, 0.81)	0.01 (-0.79, 0.84)
Exploration	-	0.19 (-0.06, 0.44)
Exploration x Linear age	-	-0.2 (-0.62, 0.22)
Exploration x Quadratic age	-	0.06 (-0.05, 0.17)
Exploration x First age	-	0.1 (-0.26, 0.47)
Exploration x Last age	-	-0.11 (-0.29, 0.06)
Random effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual	0 (0, 0)	0 (0, 0)
Plot x Year	0.56 (0.25, 1)	0.57 (0.26, 1.02)
Plot	0.37 (0.14, 0.69)	0.37 (0.14, 0.71)
Year	0.46 (0.34, 0.61)	0.47 (0.34, 0.62)
Residual	3.29 (3.29, 3.29)	3.29 (3.29, 3.29)
Adjusted Repeatability	***r* (95% CI)**	***r* (95% CI)**
Individual	0 (0, 0)	0 (0, 0)
Plot x Year	0.12 (0.06, 0.2)	0.12 (0.06, 0.2)
Plot	0.08 (0.03, 0.14)	0.08 (0.03, 0.14)
Year	0.1 (0.07, 0.13)	0.1 (0.07, 0.13)
Residual	0.71 (0.63, 0.78)	0.7 (0.62, 0.77)
Sample sizes	**n**	**n**
PlotYear	84	84
Plot	12	12
Year	7	7
Individual	813	791
Observations	1209	1187

[1]First-year breeders (i.e., second-year birds) were given age equal to zero; brood size manipulation category (BSM) was fitted as an additional fixed-effect factor with "not manipulated" as the reference (Appendix S3). The statistical intercept is therefore for second-year birds and nests that were not manipulated.

Appendix S3

As part of preliminary data analyses, we considered that effects of age of last observed reproduction could differ between birds with complete life-histories (birds not observed for two consecutive years following their last observed productive event, whether due to mortality or permanent emigration) and birds whose life-histories were potentially not yet complete (all other birds). Our primary concern was that inclusion of birds with incomplete life-histories could bias our estimates of i) linear and quadratic age effects, ii) pre-peak and post-peak age effects, or iii) the age of peak reproduction. We therefore expanded all base models by including two further fixed effects: the binary variable "complete" (code "0" for no and "1" for yes) and its interaction with age of last observed reproduction. While interactive effects between "complete" and age of last observed reproduction were supported for some traits (Table S3a), importantly, estimates or levels of support for effects of linear age, quadratic age, pre-peak age, post-peak age were not affected, and neither were estimated ages of peak reproduction (compare estimates and 95% credible intervals with Table 1). As this variable did not affect parameters of key interest, we pragmatically decided to not model (interactive) effects of complete versus incomplete life-histories in the models presented in the Main Text.

Along the same lines, we also considered whether our estimates of i) linear and quadratic age effects, ii) pre-peak and post-peak age effects, or iii) the age of peak reproduction could be biased by not considering variation in females body mass. We therefore also expanded our main models presented in Table 1 by including body mass. Though body mass affected some reproductive traits (Table S3b), as above, this variable also did not affect parameters of key interest, and we again pragmatically decided to not model these effects in the models presented in the Main Text.

Table S3a. Sources of variation in reproductive traits and exploratory activity. Analyses expand Table 1 by including controlling for completeness of life-history (coded "0" for no and "1" for yes; defined in Appendix S2) and its interaction with last age. See Table 1 for further explanations.

	Clutch Size	No. Fledglings	Fledgling Mass	Exploratory activity
	Count	Count	Grams	Count (No. hops)
Fixed effects	**β (95 CI)**	**β (95 CI)**	**β (95 CI)**	**β (95 CI)**
Intercept	8.03 (7.59, 8.48)	4.82 (3.96, 5.7)	14.89 (14.17, 15.6)	69.07 (64.26, 74.02)
Linear age	0.31 (0.12, 0.51)	0.36 (0.01, 0.71)	0.04 (-0.27, 0.35)	-3.81 (-7.01, -0.48)
Quadratic age	-0.08 (-0.12, -0.03)	-0.16 (-0.25, -0.07)	-0.02 (-0.09, 0.06)	-0.1 (-0.94, 0.71)
First age	-0.15 (-0.41, 0.11)	0 (-0.34, 0.34)	0.22 (-0.08, 0.52)	5.82 (2.34, 9.33)
Last age[1]	-0.03 (-0.21, 0.13)	0.28 (0.05, 0.49)	0.11 (-0.09, 0.31)	1.1 (-1.06, 3.25)
BSM: control (0)	NA	0.09 (-0.54, 0.71)	-0.4 (-0.95, 0.15)	2.21 (-3.91, 8.15)
BSM: enlarged (+3)	NA	1.39 (0.76, 2.03)	-0.61 (-1.13, -0.08)	-2.17 (-7.83, 3.54)
BSM: reduced (-3)	NA	-1.37 (-2.01, -0.74)	-0.5 (-1.02, 0.04)	0.91 (-5.09, 6.84)
Complete	0.16 (-0.3, 0.59)	0.86 (0.18, 1.51)	0.39 (-0.2, 0.98)	1.97 (-3.25, 7.27)
Complete x Last age[2]	0.11 (-0.09, 0.31)	-0.28 (-0.49, -0.07)	-0.14 (-0.35, 0.06)	-2.34 (-4.81, 0.08)
Random effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual	1.47 (1.36, 1.6)	0.18 (0.16, 0.21)	0.44 (0.38, 0.49)	143.24 (128.53, 158.6)
Plot x Year	0.04 (0.03, 0.05)	0.24 (0.17, 0.32)	0.33 (0.24, 0.43)	1.08 (0.76, 1.45)
Plot	0.06 (0.03, 0.11)	0.26 (0.1, 0.48)	0.21 (0.09, 0.38)	14.18 (5.63, 26.38)
Year	0.16 (0.08, 0.26)	0.93 (0.42, 1.74)	0.55 (0.26, 0.99)	10.97 (3.64, 23.72)
Residual	0.83 (0.76, 0.9)	3.04 (2.77, 3.34)	2.23 (2.03, 2.43)	274.95 (253.34, 299.3)
Adjusted Repeatability	***r* (95% CI)**	***r* (95% CI)**	***r* (95% CI)**	***r* (95% CI)**
Individual	0.58 (0.54, 0.6)	0.04 (0.03, 0.05)	0.12 (0.1, 0.13)	0.32 (0.3, 0.35)
Plot x Year	0.02 (0.01, 0.02)	0.05 (0.04, 0.07)	0.09 (0.07, 0.11)	0 (0, 0)
Plot	0.02 (0.01, 0.04)	0.05 (0.02, 0.1)	0.06 (0.03, 0.1)	0.03 (0.01, 0.06)
Year	0.06 (0.03, 0.1)	0.2 (0.1, 0.32)	0.14 (0.08, 0.24)	0.02 (0.01, 0.05)
Residual	0.32 (0.3, 0.35)	0.66 (0.55, 0.74)	0.59 (0.52, 0.65)	0.62 (0.59, 0.65)
Peak performance	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**
Trait value at peak[2]	8.36 (7.79, 8.97)	5.04 (4.1, 6.03)	NA	NA
Age at peak[3]	3.1 (2.24, 4.42)	2.06 (1.05, 2.81)	NA	NA
Pre/post-peak analysis	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**
Pre-peak age effect	0.2 (0.07, 0.32)	0.36 (0, 0.72)	NA	NA
Post-peak age effect	-0.24 (-0.6, 0.12)	-0.52 (-0.91, -0.14)	NA	NA
Sample sizes	**n**	**n**	**n**	**n**
PlotYear	84	84	84	84
Plot	12	12	12	12
Year	7	7	7	7
Individual	813	625	671	791
Observations	1209	894	962	1154

[1]Estimate is for birds with incomplete life-histories (reference category)

[2]Estimate is for birds with complete life-histories expressed as a deviation from the last age effect characterizing birds with incomplete life-histories.

Table S3b. Sources of variation in reproductive traits. Analyses expand Table 1 by controlling for female body mass (covariate). See Table 1 for further explanations.

	Clutch Size	No. Fledglings	Fledgling Mass	Exploratory activity
	Count	Count	Grams	Count (No. hops)
Fixed effects	**β (95 CI)**	**β (95 CI)**	**β (95 CI)**	**β (95 CI)**
Intercept	6.97 (5.26, 8.62)	4.86 (2.34, 7.41)	12.13 (9.91, 14.3)	53.38 (29.55, 77.7)
Linear age	0.37 (0.19, 0.56)	0.45 (0.1, 0.81)	0.02 (-0.28, 0.32)	-3.84 (-6.96, -0.61)
Quadratic age	-0.08 (-0.13, -0.03)	-0.15 (-0.24, -0.06)	0 (-0.07, 0.08)	-0.03 (-0.84, 0.75)
First age	-0.16 (-0.42, 0.11)	-0.06 (-0.4, 0.28)	0.18 (-0.14, 0.49)	5.61 (2.02, 9.12)
Last age	-0.02 (-0.14, 0.09)	0.02 (-0.12, 0.17)	0 (-0.13, 0.13)	-0.13 (-1.66, 1.38)
BSM: control (0)	NA	0.17 (-0.47, 0.81)	-0.41 (-0.97, 0.16)	2.35 (-3.46, 8.41)
BSM: enlarged (+3)	NA	1.47 (0.84, 2.11)	-0.62 (-1.14, -0.09)	-1.86 (-7.71, 3.74)
BSM: reduced (-3)	NA	-1.31 (-1.93, -0.65)	-0.51 (-1.04, 0.04)	0.71 (-5.18, 6.8)
Body mass	0.07 (-0.03, 0.17)	0.03 (-0.11, 0.17)	0.18 (0.05, 0.3)	0.98 (-0.41, 2.36)
Random effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual	1.49 (1.37, 1.61)	0.16 (0.14, 0.19)	0.45 (0.4, 0.51)	145.86 (131, 162.09)
Plot x Year	0.04 (0.03, 0.05)	0.21 (0.15, 0.29)	0.36 (0.26, 0.47)	0.93 (0.66, 1.26)
Plot	0.07 (0.03, 0.13)	0.3 (0.12, 0.57)	0.19 (0.08, 0.35)	10.81 (4.14, 20.79)
Year	0.14 (0.08, 0.22)	0.82 (0.36, 1.58)	0.62 (0.29, 1.13)	10.59 (3.9, 22.13)
Residual	0.83 (0.77, 0.9)	3.1 (2.82, 3.4)	2.19 (2, 2.39)	272.9 (250.65, 296.28)
Adjusted Repeatability	***r* (95% CI)**	***r* (95% CI)**	***r* (95% CI)**	***r* (95% CI)**
Individual	0.58 (0.55, 0.61)	0.04 (0.03, 0.04)	0.12 (0.1, 0.14)	0.33 (0.31, 0.35)
Plot x Year	0.02 (0.01, 0.02)	0.05 (0.03, 0.06)	0.09 (0.07, 0.12)	0 (0, 0)
Plot	0.03 (0.01, 0.05)	0.06 (0.03, 0.12)	0.05 (0.02, 0.09)	0.02 (0.01, 0.05)
Year	0.05 (0.03, 0.08)	0.18 (0.09, 0.3)	0.16 (0.08, 0.27)	0.02 (0.01, 0.05)
Residual	0.32 (0.3, 0.35)	0.68 (0.58, 0.76)	0.58 (0.5, 0.64)	0.62 (0.59, 0.64)
Peak performance	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**
Trait value at peak[2]	7.4 (5.63, 9.11)	5.22 (2.65, 7.82)	NA	NA
Age at peak[3]	3.35 (2.67, 4.49)	2.51 (1.62, 3.29)	NA	NA
Pre/post-peak analysis	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**
Pre-peak age effect	0.2 (0.08, 0.33)	0.41 (0.05, 0.77)	NA	NA
Post-peak age effect	-0.25 (-0.63, 0.13)	-0.5 (-0.89, -0.1)	NA	NA
Sample sizes	**n**	**n**	**n**	**n**
PlotYear	84	84	84	84
Plot	12	12	12	12
Year	7	7	7	7
Individual	801	620	663	783
Observations	1179	880	944	1137

Appendix S4

We controlled for the effects of brood size manipulations (conducted in 2010 and 2011) by including a four-level fixed-effect factor for all traits except those quantified prior to the brood size manipulation (clutch size). This factor was coded as "0" = not manipulated (n = 1000 nests) vs "1" = control (n = 66 nests) vs "2" = reduced (n = 71 nests) vs "3" = enlarged (n = 72 nests). Category "not manipulated" was set as the reference, such that the statistical intercept estimates printed in Tables 1 and 2 represent the mean value for nests that were not manipulated. The effects of these brood size manipulations are extensively detailed elsewhere (Nicolaus et al. 2015) and therefore not presented in the Main Text. Briefly, (i) the number of fledglings (in broods fledging any offspring) did not differ between control broods and broods that were not manipulated, but enlarged (reduced) broods fledged fewer (more) nestlings then nests that were not manipulated, (ii) the average fledging mass was decreased for the enlarged brood size manipulation category, and neither the (iii) binary probability to produce any fledglings (Supplementary Table S1) nor (iv) exploratory activity differed between broods that were manipulated versus not manipulated, in line with analyses of these data published previously (Nicolaus et al. 2015).

Appendix S5

We investigated whether estimates of linear and quadratic age, as well as estimates of the age of peak performance (printed in Table 1) were biased by the inclusion of age categories with fewer than five data points; this applied to age=7 which was represented only by one data point. We therefore re-ran our main analyses (Table 1) after combining this category with the previous one (age=6). Estimates of linear or quadratic age did not change, and neither did estimates of the age of peak performance (compare Supplementary Table S5 and Table 1). We therefore concluded that the inclusion of the age category with few data points did not bias our estimates.

Table S5. Sources of variation in reproductive traits (clutch size, number and average mass of fledglings (for nests producing any fledglings)) and exploratory activity for models with the identical fixed and random effects structures are detailed in Table 1 but where age categories 6 and 7 were lumped prior to analyses.

	Clutch Size Count	No. Fledglings Count	Fledgling Mass Grams	Exploratory activity Count (No. hops)
Fixed effects	**β (95 CI)**	**β (95 CI)**	**β (95 CI)**	**β (95 CI)**
Intercept[1]	8.14 (7.8, 8.48)	5.37 (4.69, 6.07)	15.13 (14.5, 15.76)	70.13 (66.47, 73.81)
Linear age	0.35 (0.17, 0.54)	0.41 (0.06, 0.75)	0.07 (-0.22, 0.36)	-3.94 (-7.1, -0.75)
Quadratic age	-0.08 (-0.12, -0.03)	-0.14 (-0.23, -0.05)	-0.01 (-0.08, 0.07)	-0.04 (-0.86, 0.76)
First age	-0.13 (-0.4, 0.13)	-0.02 (-0.36, 0.32)	0.2 (-0.1, 0.51)	5.59 (2.08, 9.15)
Last age	-0.02 (-0.14, 0.09)	0.04 (-0.1, 0.18)	0 (-0.13, 0.13)	-0.03 (-1.53, 1.48)
BSM: control (0)	NA	0.11 (-0.52, 0.74)	-0.38 (-0.93, 0.18)	2.27 (-3.62, 8.28)
BSM: enlarged (+3)	NA	1.42 (0.8, 2.06)	-0.58 (-1.12, -0.03)	-1.88 (-7.39, 3.9)
BSM: reduced (-3)	NA	-1.35 (-1.99, -0.73)	-0.48 (-1.03, 0.06)	0.88 (-5.05, 6.67)
Random effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual	1.48 (1.36, 1.61)	0.23 (0.19, 0.26)	0.43 (0.37, 0.48)	143.74 (129.12, 159.32)
Plot x Year	0.04 (0.03, 0.06)	0.26 (0.19, 0.35)	0.33 (0.24, 0.43)	0.77 (0.55, 1.04)
Plot	0.07 (0.03, 0.12)	0.24 (0.09, 0.45)	0.21 (0.09, 0.39)	13.52 (5.22, 25.6)
Year	0.15 (0.08, 0.24)	0.77 (0.33, 1.45)	0.59 (0.28, 1.1)	11.15 (3.85, 24.07)
Residual	0.83 (0.76, 0.89)	3.02 (2.76, 3.31)	2.23 (2.04, 2.45)	276.18 (255.24, 299.43)
Adjusted Repeatability	***r* (95% CI)**	***r* (95% CI)**	***r* (95% CI)**	***r* (95% CI)**
Individual	0.58 (0.55, 0.61)	0.05 (0.04, 0.06)	0.11 (0.1, 0.13)	0.32 (0.3, 0.35)
Plot x Year	0.02 (0.01, 0.02)	0.06 (0.04, 0.08)	0.09 (0.06, 0.11)	0 (0, 0)
Plot	0.03 (0.01, 0.05)	0.05 (0.02, 0.1)	0.06 (0.02, 0.1)	0.03 (0.01, 0.06)
Year	0.06 (0.03, 0.09)	0.17 (0.08, 0.28)	0.15 (0.08, 0.26)	0.02 (0.01, 0.05)
Residual	0.32 (0.3, 0.35)	0.67 (0.58, 0.75)	0.59 (0.52, 0.65)	0.62 (0.59, 0.65)
Peak performance	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**
Trait value at peak[2]	8.57 (8.13, 8.99)	5.69 (4.94, 6.47)	NA	NA
Age at peak[3]	3.5 (2.67, 4.7)	2.4 (1.44, 3.16)	NA	NA
Pre/post-peak analysis	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**
Pre-peak age effect[4]	0.2 (0.07, 0.32)	0.36 (0.01, 0.71)	NA	NA
Post-peak age effect[5]	-0.24 (-0.6, 0.12)	-0.52 (-0.9, -0.14)	NA	NA
Sample sizes	**n**	**n**	**n**	**n**
PlotYear	84	84	84	84
Plot	12	12	12	12
Year	7	7	7	7
Individual	813	625	671	791
Observations	1209	894	962	1154

Appendix S6

The use of BLUPS in analysis has been criticized because most studies doing so fail to account for the uncertainty in BLUP estimates (Hadfield et al. 2010; Houslay & Wilson 2017). A proposed solution is to estimate a posterior distribution of possible BLUP values and perform the subsequent analysis using this posterior distribution. Using simulations, detailed below, we show that while this approach may provide information on how the uncertainty in the BLUP estimates propagates in to the subsequent analysis, the mean estimate of the relationship becomes underestimated (i.e. biased).

This insight is derived from a simulation were the among-individual variation in a behavioural trait was set to 0.3. In other words, the variance among the intrinsic values (i.e., "personality") of different individuals was 0.3. An individual's intrinsic behavioural value was then set to affect a life history trait, were an increase in the intrinsic behavioural value increased the life history trait with a value of 0.5. We subsequently simulated several scenarios were the within-individual variation in the behavioural trait ranged from 0.001 to 1 in increments of 0.1. We simulated data sets with 100 individuals and 4 repeated measures per individual. The uncertainty in the BLUP estimates thus differed between simulations, increasing with increasing values of within-individual behavioural variance. We simulated 100 realizations for each scenario. We then used the two-step approach detailed in the Main Text. We first estimated each individual's BLUP using a mixed effects model with individual identity fitted as a random effect, and calculated a posterior distribution of the individual BLUPs using approaches and software packages detailed in the Main Text. Second, we assessed the effect of the behavioural BLUP on the simulated life history trait using either each individual's most likely BLUP estimate as input (as we do in the Main Text), or by running the analysis over the whole posterior distribution of BLUP values (i.e., 1000 times). Finally, we estimated the mean and 95% credible intervals for the bias in the estimated relationship for each of the two scenarios. The R-code (called code.R) for these simulations is attached as supplementary information.

The simulations demonstrated that the estimated mean relationship between the life-history trait and behavioural BLUP was underestimated when based on the whole posterior distribution (Figure S6a). The magnitude of the downward bias increased with increasing uncertainty (i.e., within-individual variance) in BLUP estimates (Figure S6a). By contrast, using only the most likely BLUP estimate to quantify the relationship between an individual's intrinsic value and

the life history trait did not produce biased estimates (Figure S6b) though its precision decreased with increasing within-individual variance in the BLUP. We preferred this latter approach because while estimates were more uncertain they were not biased.

Figure S6a. Bias in the estimated effect of behavioural BLUP on a simulated life-history trait as a function of the amount of simulated within-individual variation in behaviour for analyses taking forward the posterior distribution of individual BLUP values.

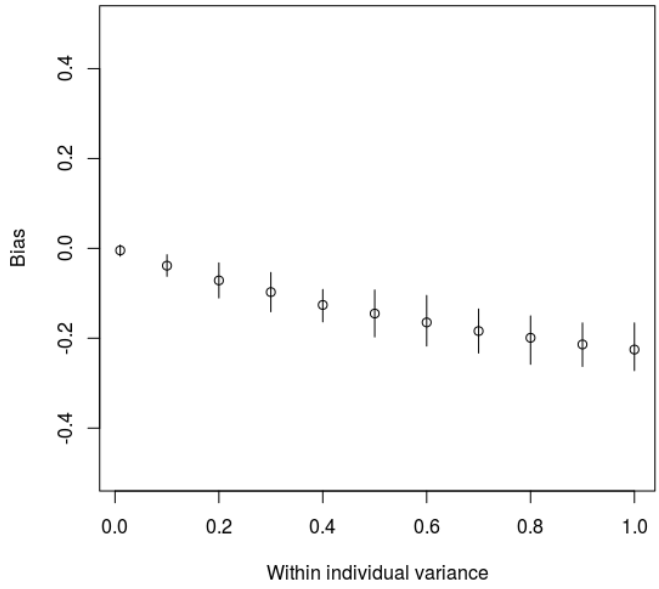
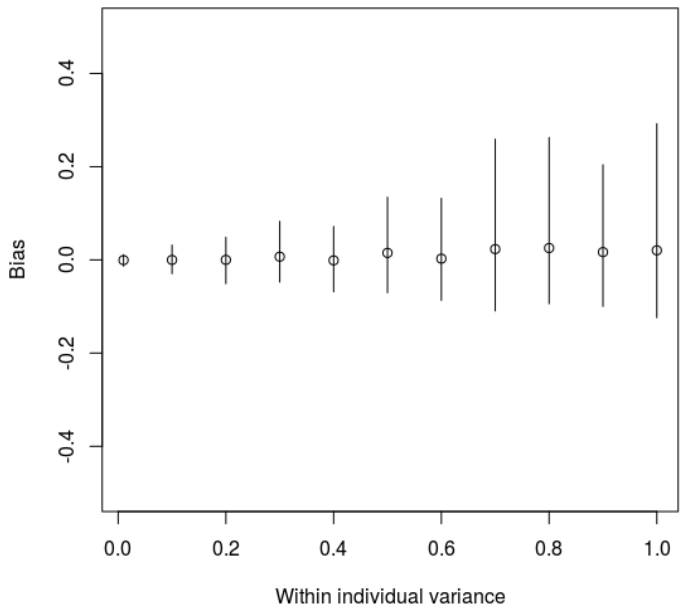


Figure S6b. Bias in the estimated effect of behavioural BLUP on a simulated life-history trait as a function of the amount of simulated within-individual variation in behaviour for analyses using an individual's mean BLUP value as input.



Appendix S7

The distribution of age classes differed between the 50% slowest (average values < 0) and 50% fastest (values ≥ 0) explorers. The slowest explorers only had reproductive data for one- to five-year olds (ages 1-5: n = 293, 187, 73, 28, 8 first clutches), while the fastest explorers had data for one- to seven-year olds (ages 1-7: n = 295, 182, 77, 30, 8, 5, 1). A follow-up analysis using only age categories where both types had data (i.e. excluding data for six- or seven-year old fast explorers; n = 6) resulted in the same level of support for interactive effects between exploratory activity and linear and quadratic age (compared analyses of clutch size in Supplementary Table S7 versus Table 2). Thus, our finding of personality-related age-dependency of clutch size was not an artefact caused by lack of data for older slow explorers.

Table S7. Sources of variation in clutch size. Analyses are identical to those presented in Table 2 though data for six- or seven-year old fast explorers (n = 6) are excluded. See Table 2 for further explanations.

	Clutch Size Count
Fixed effects	**β (95 CI)**
Intercept	8.16 (7.82, 8.5)
Linear age	0.31 (0.1, 0.52)
Quadratic age	-0.06 (-0.12, 0)
First age	-0.12 (-0.38, 0.15)
Last age	-0.03 (-0.15, 0.09)
BSM: control (0)	NA
BSM: enlarged (+3)	NA
BSM: reduced (-3)	NA
Exploration	-0.04 (-0.21, 0.11)
Exploration x Linear age	0.18 (0, 0.35)
Exploration x Quadratic age	-0.06 (-0.11, -0.01)
Exploration x First age	-0.42 (-0.69, -0.15)
Exploration x Last age	0.08 (-0.02, 0.18)
Random effects	**σ^2 (95% CI)**
Individual	1.48 (1.37, 1.61)
Plot x Year	0.04 (0.03, 0.05)
Plot	0.07 (0.03, 0.13)
Year	0.14 (0.08, 0.25)
Residual	0.83 (0.76, 0.9)
Adjusted Repeatability	***r* (95% CI)**
Individual	0.58 (0.54, 0.61)
Plot x Year	0.02 (0.01, 0.02)
Plot	0.03 (0.01, 0.05)
Year	0.06 (0.03, 0.09)
Residual	0.32 (0.3, 0.35)
Sample sizes	**n**
PlotYear	84
Plot	12
Year	7
Individual	791
Observations	1181

Appendix S8

Exploratory behaviour exhibited a "first observed age" effects suggestive of selective appearance. A follow-up analyses demonstrated that this effect could be attributed fully to individual differences in experience. Specifically, we calculated a parameter representing the estimate of "linear age" minus the estimate of "first observed age"; this parameter did not deviate from zero (mean, 95 CIs: 1.59, -1.83, 4.99). This finding implied that there was no statistical evidence for selective appearance with respect to individual-specific age-related reaction norm intercepts for exploratory activity when biasing effects of within-individual plasticity were taken into account. Briefly, the decrease in exploratory behaviour with linear age (Table 1), representing an effect of experience with the assay (Dingemanse et al. 2012), caused birds recruiting into the breeding population at an older age to behave faster.

SI REFERENCES

Dingemanse, N.J., Bouwman, K.M., van de Pol, M., van Overveld, T., Patrick, S.C., Matthysen, E. & Quinn, J.L. (2012) Variation in personality and behavioural plasticity across four populations of the great tit Parus major. Journal of Animal Ecology, 81, 116-126.

Hadfield, J.D., Wilson, A.J., Garant, D., Sheldon, B.C. & Kruuk, L.E.B. (2010) The misuse of BLUP in ecology and evolution. American Naturalist, 175, 116-125

Houslay, T.M. & Wilson, A.J. (2017) Avoiding the misuse in BLUP in behavioral ecology. Behavioral Ecology, 28, 948-952.

Nicolaus, M., Mathot, K.J., Araya-Ajoy, Y.G., Mutzel, A., Wijmenga, J.J., Kempenaers, B. & Dingemanse, N.J. (2015) Does coping style predict optimization: an experimental test in a wild passerine bird. Proceedings of the Royal Society B-Biological Sciences, 282, 20142405

Chapter 2

Heterogeneous selection on exploration behavior within and among West European populations of a passerine bird

Published in Proceedings of the National Academy of Sciences of the United States of America (2021) **118** (27)

Heterogeneous selection on exploration behavior within and among West European populations of a passerine bird

Alexia Mouchet, Ella F. Cole, Erik Matthysen, Marion Nicolaus, John L. Quinn, Allison M. Roth, Joost M. Tinbergen, Kees van Oers, Thijs van Overveld, Niels J. Dingemanse

Abstract

Heterogeneous selection is often proposed as a key mechanism maintaining repeatable behavioral variation ("animal personality") in wild populations. Previous studies largely focused on temporal variation in selection within single populations. The relative importance of spatial versus temporal variation remains unexplored, despite these processes having distinct effects on local adaptation. Using data from >3500 great tits (Parus major) and 35 nest box plots situated within five West-European populations monitored over 4-18 years, we show that selection on exploration behavior varies primarily spatially, across populations, and study plots within populations. Exploration was, simultaneously, selectively neutral in the average population and year. These findings imply that spatial variation in selection may represent a primary mechanism maintaining animal personalities, likely promoting the evolution of local adaptation, phenotype-dependent dispersal, and nonrandom settlement. Selection also varied within populations among years, which may counteract local adaptation. Our study underlines the importance of combining multiple spatiotemporal scales in the study of behavioral adaptation.

Keywords: animal personality, macro-spatial variation, fluctuating selection, integrative fitness, local adaptation

INTRODUCTION

Repeatable and heritable variation in behavior ("animal personality") is ubiquitous among wild animal populations (1). Repeatable behavioral differences among individuals can be adaptive when the costs and benefits of alternative behavioral tactics vary with the environment (2, 3). This requires heterogeneous selection, either spatially, temporally, or spatiotemporally, e.g., within or among populations, habitats or years (3, 4). Social environments may also play a key role by inducing negative frequency-dependent selection (5, 6). Heterogeneous selection on repeatable individual variation in behavior has previously been demonstrated primarily within single populations sampled over limited numbers of years (7, 8), thus limiting our understanding of the relative importance of spatial and temporal variation in selection.

Spatial and temporal processes co-occur (8, 9) but have distinct effects on population dynamics and evolution. Strong spatial variation favors different behavioral phenotypes in different locations, which may induce selection for nonrandom dispersal, and rapid population divergence (10, 11). Temporal variation instead favors the coexistence of multiple behavioral phenotypes within populations, thereby counteracting population divergence. Estimates of selection from multiple study populations monitored over multiple years are required to estimate spatial and temporal variation simultaneously, and determine their relative importance in maintaining individual behavioral variation.

To address this question, we assayed exploration behavior in a novel environment (12) among great tits (Parus major) breeding in 35 nest box plots across five populations in Western Europe, each sampled for multiple (4-18) years. For four of these populations, animal model-based quantitative genetics were conducted; in all cases exploration behavior was significantly repeatable and heritable (Boshoek, Belgium: $R=0.42$, $h^2=0.30$ (13, 14); Lauwersmeer, the Netherlands: $R=0.40$-0.44, $h^2=0.10$-0.11 (13, 15); Westerheide, the Netherlands: $R=0.38$, $h^2=0.14$ (13, 16); Wytham Woods, United Kingdom: $R=0.34$, $h^2=0.26$ (13, 16). We estimated the average pattern of selection (directional, stabilizing, disruptive) within the average population, plot, and year, and examined whether selection was heterogeneous as predicted by state-dependent personality models (3, 4). Finally, we estimated the relative proportion of variation in selection that was attributable to five distinct sources: a) macro-spatial variation (among populations), b) micro-spatial variation (among plots within populations), c) temporal variation (among years), d) population-specific (or macro-scale) temporal variation (unique combinations of population and year), and e) plot-specific (or micro-scale) temporal variation (unique combinations of plot and year) (Fig. 1).

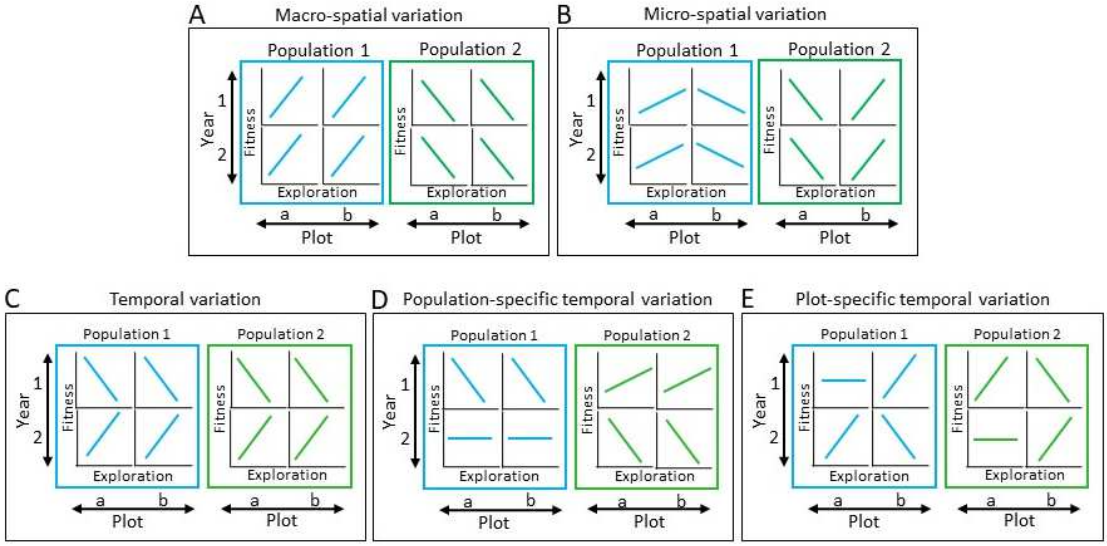


Figure 1. Distinct heterogeneous selection scenarios illustrated for two populations with two plots each sampled over the same two years. Selection can vary A) macro-spatially (among populations), B) micro-spatially (among plots), C) temporally (among years), D) macro-spatiotemporally (year-effects are population-specific) and E) micro-spatiotemporally (year-effects are plot-specific). Illustrated scenarios are mutually nonexclusive; our analyses of the sources of variation in selection on exploration behavior imply all mechanisms are important.

RESULTS

Exploration behavior was neither subject to linear nor to nonlinear viability (adult survival) selection within the average population, plot and year, though estimates of nonlinear selection suggested weak disruptive viability selection (γ, Table 1; SI Appendix Table S1). By contrast, exploration behavior was, on average, subject to stabilizing local offspring recruitment selection (γ, Table 1; SI Appendix Table S1). The effect of stabilizing local recruitment selection appeared to be cancelled out by the weak effect of disruptive viability selection. Indeed, selection measured using integrative fitness, which combines annual survival and local recruitment, was not different from zero (Table 1; SI Appendix Table S1). Previous research has shown that faster explorers disperse further (14, 17). Consistent with this idea, immigrants are often faster explorers than local recruits (14, 17, 18); this was also the case in our dataset (SI Appendix Table S2). We therefore examined whether our estimates of local recruitment selection were biased against faster explorers. To do so, we re-estimated fecundity (and integrative) fitness selection using annual fledgling production, a pre-dispersal reproductive success metric that should not suffer from dispersal-related bias. These analyses produced the same results as reported above, thus suggesting that nonrandom dispersal did not bias our estimates of selection (SI Appendix Table S3). We conclude, therefore, that exploration behavior was indeed selectively neutral overall.

Table 1. Linear (β) and nonlinear (γ) standardized selection gradients estimated for exploration behavior, with integrative fitness, adult survival (viability) or local offspring recruitment as fitness metrics. Estimates, with 95% credible interval (CI), are derived from random regression models fitting exploration behavior standardized over the entire dataset.

	Integrative fitness	**Survival**	**Local recruitment**
Selection gradient	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)
β	0.02 (-0.45, 0.48)	0.02 (-0.49, 0.52)	0.06 (-0.56, 0.67)
γ	0.00 (-0.03, 0.02)	0.02 (-0.02, 0.05)	-0.10 (-0.11, 0.01)

Exploration behavior was, however, under heterogeneous selection. This conclusion was supported for all fitness metrics considered based on hierarchical random regression analyses and associated permutation tests (Table 2; SI Appendix Table S1). Combined with evidence for neutral selection overall, our finding of heterogeneous directional selection implies that selection varied not just in strength but also in direction (Table 2, Fig.2). The effect of exploration on integrative fitness varied spatially and temporally at both macro- and micro-scales (variance in random slopes; all permutation $P<0.01$, SI Appendix Table S1). Macro-spatial variation in selection explained the largest percentage of the total variance in selection (effect of population; 47%; Table 2). Micro-spatial (plot; 13%), temporal (year: 19%), population-specific (i.e., macro-scale) temporal (population × year; 11%) and plot-specific (i.e., micro-scale) temporal (plot × year; 9%) variation in selection existed but were of lesser importance (Table 2). Analyses of survival and local recruitment led to the same conclusions: macro-spatial variation in selection explained the largest percentage of the variance in selection (viability selection: 35%; recruitment selection; 39%); variation in selection at other spatiotemporal levels explained similar relative amounts of variance as described for integrative fitness (Table 2, SI Appendix Table S1). This implies that overall patterns of variation in selection were similar for both fitness components.

Table 2. Proportion of variance in selection attributable to each ecological level with associated 95% credible intervals (CIs), for integrative fitness, adult survival (viability) or local offspring recruitment as focal fitness metric.

	Integrative fitness	**Survival**	**Local recruitment**
Ecological level	R (95% CI)	R (95% CI)	R (95% CI)
Population	0.47 (0.37, 0.60)	0.36 (0.28, 0.46)	0.39 (0.31, 0.50)
Plot	0.13 (0.10, 0.16)	0.16 (0.13, 0.17)	0.16 (0.13, 0.19)
Year	0.19 (0.16, 0.21)	0.17 (0.15, 0.18)	0.18 (0.16, 0.19)
Population x Year	0.11 (0.08, 0.14)	0.13 (0.11, 0.16)	0.13 (0.10, 0.16)
Plot x Year	0.09 (0.06, 0.12)	0.18 (0.15, 0.20)	0.13 (0.10, 0.16)

Statistical support for heterogeneous selection can occur as an artefact when phenotypes vary among levels of random effects in situations where selection is nonlinear (SI Appendix Fig. S1). We addressed this concern by re-fitting our models to incorporate level-specific nonlinear patterns of selection, which were not detected (SI Appendix Table S4), thus suggesting that this concern was unfounded. Patterns of (variation in) selection also did not depend on whether trait values were standardized at the lowest hierarchical level (i.e. within unique combinations of plot and year) rather than over the entire dataset (SI Appendix Table S5).

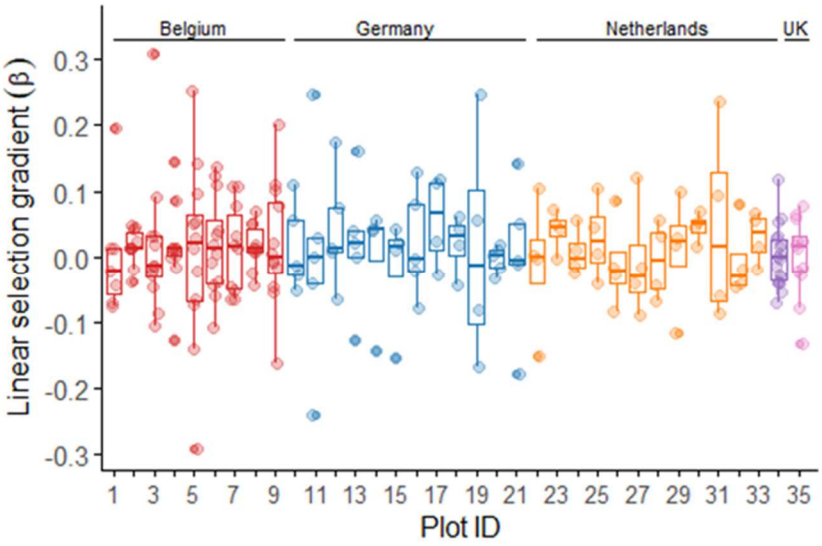


Figure 2. Patterns of heterogeneous selection on exploration behavior within and among five great tit populations sampled across Western Europe. Colors represent populations, which were located across four countries. Boxplots show the median, first and third quartile of the standardized selection gradient (with whiskers) for each study plot, and dots the standardized selection gradient for each sampled year within a focal plot. While some populations had multiple plots (red: Boshoek, Belgium; blue: Starnberg, Germany; orange: Lauwersmeer, the Netherlands), other populations consisted of a single plot (purple: Westerheide, the Netherlands; pink: Wytham Woods, UK). We used integrative fitness as our metric of annual fitness. Positive (vs. negative) selection gradients indicate selection favoring fast (vs. slow) explorers.

DISCUSSION

Strong spatial variation in selection can induce selection for individuals to settle in habitats best fitting their behavioral phenotype (19). Temporal variation would instead favor the coexistence of multiple behavioral phenotypes within populations. Our analyses showed that selection on exploration behavior varied macro-spatially, micro-spatially and temporally. We further detected evidence for population- and plot-specific differences in patterns of temporal variation in selection. Additionally, integrative fitness selection was neutral overall due to weak disruptive viability selection counterbalanced by stabilizing local recruitment selection. Our

findings imply that heterogeneous selection on personality exists at many (if not all) major ecological levels, and, in combination with neutral selection overall, that environmental variation at multiple spatial and temporal scales contributes to the maintenance of behavioral variation.

Our discovery of a major role for macro-spatial variation in selection implies that large-scale geographical variation in ecological factors has the potential to select for population divergence, which, consequently, might promote nonrandom dispersal and settlement. Specifically, individuals should settle in habitats where they do best, which will differ between behavioral phenotypes (19, 20). Ecological factors that constrain dispersal of certain genotypes (e.g., winter temperature) may also facilitate genetic differentiation (21). Moreover, behavioral phenotypes may choose populations based on social rather than nonsocial environmental conditions. For example, positive frequency-dependent selection favors non-aggressive Western bluebirds (Sialia mexicana) when surrounded by non-aggressive conspecifics because those are more cooperative (22). Similarly, fast-exploring great tits are known to acquire relatively low dominance ranks as first-year birds, and are more likely to disperse away from their natal area (17, 23); this may result in fast-exploring birds consequently settling in less competitive (i.e., low density) areas where their behavioral phenotype may perform best (24), and supports the nonrandom-dispersal hypothesis. Regardless of the causal factors, nonrandom dispersal may thereby reinforce assortative mating (25), induce biased gene flow, accelerate (genetic) population divergence, and eventually, enhance population evolvability (26). The simultaneous occurrence of micro-spatial variation in selection (i.e., among plots within populations) implies that selection also favors local adaptation among habitats of the same population. Local adaptation within populations may, however, often be counteracted by substantial gene flow given the species' dispersal characteristics (21, 27). Studies characterizing selection on personality-dependent habitat choice (a form of phenotype-environment matching (19)) and dispersal are required to reveal the interplay of mechanisms shaping evolutionary trajectories of behavioral traits in natural populations.

Temporal fluctuations favors certain behavioral phenotypes in certain years and other phenotypes in other years (3). Fluctuating selection will thus inherently counteract the speed of population divergence and consequently, plays a key role in preventing genetic differentiation required in the process of local adaptation. The existence of behavioral variation in all five great tit populations despite the large magnitude of macro-scale spatial variation in selection however suggests that local adaptation may be reduced by the combined action of temporal variation at small spatial scales and of gene flow at larger spatial scales. This may

explain why the combined additive and interactive effects of plot, year, and population explained as little as 4% of the variance in exploration behavior among first-year birds (SI Appendix Table S2). A key question is therefore at what spatial scale which mechanism predominantly counteracts population divergence and whether genetic population divergence in behavior occurs at all. Forcefully addressing this question would require study plots at spatial scales intermediate to our within- and among-population levels, e.g., multiple populations within countries.

Temporal variation in selection can also result in the evolution of reversible plasticity but previous great tit studies suggest that limits to plasticity prevent this mechanism from evolving (28). Temporal variation in selection resulted from ecological factors varying over large spatial scales, but also from local fluctuations. Specifically, our finding of a main effect of year on selection reveals that selection on personality changes in concert across large geographical scales. These selection pressures likely result from ecological factors varying in conjunction across Western-Europe (29). Beech (Fagus sylvatica) masting, a phenomenon where beech trees produce high numbers of seeds in some years but few (or none) in other years, may represent such a key biotic factor. Beech masting strongly affects winter survival of great tits (30), and is often synchronized over the entire continent (31). Such temporal variation in food availability (and selection), however, will be evidently modulated by local habitat conditions, such as tree species composition. This may explain why we also found strong evidence for population-specific (i.e., macro-scale) temporal variation in selection (population × year effects). Other factors may also play a key role here, for example, predator- or parasite-induced selection varying more among years in populations with high versus low overall levels of these biological factors (32–35). Our finding of plot-specific (i.e., micro-scale) temporal variation in selection (plot × year effects) indicates that similar factors act among habitat patches within populations.

Previous studies revealed a key role of social environmental variation by demonstrating that selection on exploration behavior (18, 28) and aggression (22) varies with breeding density within populations. We investigated this explanation by expanding our models to include the interactive effect of breeding density (pair/ha) and exploration on fitness. Doing so did not result in a detectable change in random slope variance at any of the hierarchical levels (SI Appendix Table S6). This implies that heterogeneous selection largely resulted from ecological processes independent of density, yet to be determined. Here, social interactions inducing negative frequency-dependent selection may constitute a key mechanism contributing to the maintenance of variation (5). Forcefully investigating this idea requires large sample sizes for

each social environment (here, each unique combination of plot and year) to accurately and precisely estimate phenotype frequencies. The many small plots characterizing our study do not fulfil this requirement.

A popular explanation for the persistence of personalities is that behavioral phenotypes differ in how they resolve life-history trade-offs (36). Personality-related pace-of-life theory predicts that fast explorers produce larger clutches but either live less long, or senesce at an earlier age, compared to slow explorers (36, 37). Though previous work on great tits supports some of these predictions (38–40), meta-analyses do not (41, 42). Our analyses, similarly, fail to find support for personality-related pace-of-life syndromes: annual adult survival was not lower for faster explorers, and nor did this type of bird produce more fledglings or local recruits annually compared to slower explorers. The detected pattern of stabilizing recruitment selection combined with weak disruptive viability selection implies that other ecological explanations are required to explain any personality-related differences in life-history in this system (43, 44).

Nonrandom natal dispersal may bias estimates of variance in recruitment selection, though this would require that, in different places or at different times, different behavioral phenotypes are most dispersive; this condition is unlikely met at all five spatiotemporal scales at which we detected heterogeneous recruitment selection. Moreover, variance estimates of fecundity selection using annual fledgling number were similar to variance estimates of local recruitment selection, refuting the idea that our estimates of heterogeneous selection measured through local recruitment rates were biased (SI Appendix Table S3). Capture-mark-recapture analyses have demonstrated that capture rates do not vary with exploration behavior among adult breeders (45). As great tits show limited breeding dispersal (46, 47), sampling bias cannot easily affect the variance in adult viability selection. Altogether, these arguments suggest that sampling bias does not play a major role in explaining the spatiotemporal patterns of heterogeneous selection revealed by this study.

Our analyses of temporal and spatial patterns of variation in selection represent an important contribution to our understanding of population dynamics and the evolution of behavior. Macro-spatial variation in selection counteracted by temporal variation demonstrates the importance of estimating heterogeneous selection on individual behavior at multiple hierarchical scales. Microevolutionary responses to selection now require study to reveal whether the spatial patterns of variation in selection uncovered by this study reduce the genetic variation in behavior within populations and whether temporal variation combined with gene flow are indeed sufficient to prevent this erosion of genetic variation due to population-specific

fitness optima. Our insights are likely not specific to selection on behavior or personality but may apply generally, and warrant analyses of spatiotemporal variation in selection for other key phenotypic traits, such as physiology, morphology or life-history traits. Our study exemplifies the need for long-term studies across multiple habitats, and international collaborations to reveal large-scale geographical patterns of selection and the key role of ecology in shaping selection and evolution (48).

MATERIALS AND METHODS

Study populations and field data collection

Data were collected in five nest box populations of great tits between 2006-2017 (Boshoek near Antwerp, Belgium; 51°08'N, 043°2'E), 2006-2009 (Lauwersmeer, The Netherlands; 53°20'N, 06°12'E), 2010-2014 (Starnberg District, Bavaria; Germany; 47°58'N, 11°14'E), 1999-2016 (Westerheide; The Netherlands; 52°00'N, 05°50'E) and 2005-2016 (Wytham Woods, United Kingdom; 51°47'N, 1°20'W). In Boshoek, nine nest box plots were fitted in 0.6-9 ha woodland fragments at a density of six boxes per ha (49). In both Bavaria and Lauwersmeer, 12 nest box plots were fitted in 8-11 ha woodland fragments at a density of 4.5 to 6.2 boxes per ha (28). In Westerheide and Wytham Woods, a single nest box plot was fitted within continuous woodland habitat of, respectively, ca. 112 and 290 ha at a density of 3 and 3.5 boxes per ha.

We checked nest boxes at least once a week during the breeding season (April-July) to determine key life-history traits and breeding density. Breeders were caught in their nest box when their nestlings were 7-12 days old and ringed at first capture. We also ringed offspring before fledging to determine which offspring recruited into the population as breeders in subsequent years. Outside the breeding season, birds were captured in nest boxes when roosting (November-February; all populations except Wytham Woods) and/or with mist-nets (July-March, in Boshoek, Westerheide and Wytham Woods).

Exploration assays

We assayed exploration behavior under standardized laboratory conditions using a novel environment test (50) made suitable for wild birds (12). Prior to the test, birds were individually housed in a small cage overnight with ad libitum access to food and water. Each cage connected to the novel environment, a standard laboratory room fitted with five artificial trees, via a sliding door that allowed release without handling (12). Slight differences in setup and

procedure existed across populations as detailed elsewhere (13, 51). An exploration score was calculated by summing up the total number of flights and hops between perches made within the first two minutes after entering the room (13). This score of movement behavior genetically correlated with the number of areas visited, thus it represents a good proxy of spatial exploration (18). Birds were tested between 8h00-13h00. The dataset consisted of 5459 records collected from 3551 individuals typed for exploration behavior, distributed over 188 unique combinations of plot and year ("plot-years").

Data characterization and selection

To estimate selection on exploration behavior, we used an integrative measure of fitness that represents an individual's overall annual fitness. We calculated it as 1 × the focal adult's survival probability + 0.5 × the number of its locally recruited offspring for a given year. This integrative fitness measure acknowledges that each individual contributes fully its genes to the next year when returning as a breeder but that only half of its genes are present in any recruited offspring (52, 53). This inherently avoids biases attributable to individual differences in how trade-offs between offspring quality and quantity, or between current and future reproduction are resolved (36, 37, 42). We defined adult survival as the binary probability that a focal bird breeding in a focal year was found breeding in the following year (binomial; not found [0] or found breeding [1]). In this species, capture probabilities of breeders are high (75-95%) (45) and breeding dispersal rates low (46, 47), implying that this metric appropriately measures local survival. We defined local offspring recruitment as the annual number of offspring recruiting as breeders in the focal population (regardless of plot identity). Because nonrandom dispersal can bias estimates of fecundity selection based on counting local recruits, we also estimated selection using the annual number of produced fledglings as alternative metric.

Our dataset included only individuals for which exploration behavior was scored prior to a focal breeding season; this avoids bias in estimates of adult survival and local offspring recruitment between subsequent breeding seasons (45). We used the first exploration score of each individual as a measure of exploration behavior. We assumed this reflected an individual's personality (54) because elsewhere we show that individual-mean values (proposed to best reflect an individual's average behavior (55)) are tightly positively correlated with an individual's first exploration score among repeatedly assayed birds (28). We did not use individual-mean values because (i) many individuals were not tested repeatedly (i.e., individual-mean values would be based on unequal replication between individuals), and (ii)

individuals differ in how exploration behavior changes with repeated exposure to the testing procedure (13).

Statistical analyses

We estimated selection on exploration behavior using both our integrative fitness metric and its underlying components (adult survival and local offspring recruitment or fledgling production). Doing so enabled identifying whether selection acted via specific pathways (e.g., via survival rather than recruitment selection (28)) and whether selection varied in the same proportion at each hierarchical level for each fitness component.

We fitted generalized linear mixed-effects models (GLMMs) with Gaussian (integrative fitness analyses), binomial (logit link; survival analyses), and Poisson errors (recruitment and fledgling analyses). Each model simultaneously estimated the magnitude of variation in directional selection among populations (macro-spatial variation), plots (micro-spatial variation), years (micro-temporal variation), unique population-year combinations (population-specific or macro-scale temporal variation) and unique plot-year combinations (plot-specific or micro-scale temporal variation). This was achieved by fitting random intercepts and slopes (with respect to exploration score fitted as a fixed effect covariate (28)) at each of these hierarchical levels. Insights into variation in nonlinear selection (i.e., in shape of selection) would require fitting nonlinear random slopes, however, our data do not provide enough statistical power to forcefully address this question.

Exploration scores were corrected for seasonal plasticity (12, 13) following Ref. (12) to avoid biased estimates (56). They were then squared-root transformed and standardized (i.e., zero mean and unit standard deviation) to acquire (standardized) selection gradients. We performed this standardization over the entire dataset because this produces estimates that are comparable across all hierarchical levels (57). However, great tits experience strong density-dependent selection within plots among years (“plot-years’) (58, 59), and previous studies implied that traits should be standardized at the level at which competition occurs (60). We therefore also ran our analyses after standardizing traits within plot-years. We estimated linear and nonlinear (quadratic) selection on exploration behavior to test for directional and disruptive or stabilizing selection. Nonlinear selection was assessed by adding the squared term of the standardized exploration value (defined above) as a fixed effect covariate. Standardized linear and nonlinear selection gradients were estimated by re-running our models using relative fitness (i.e., the focal fitness metric divided by the grand mean of the dataset) as a response

variable; quadratic selection gradients were calculated by doubling the estimated parameter for the square of exploration (61).

We ran all analyses in R v. 3.5.3. (62), using the Bayesian inference package R-INLA (63) and the "iid2d" model. We estimated posterior means and their 95% credible intervals (CIs) for all fixed and random effects. Fixed effect priors were normally distributed with zero mean and precision (inverse of variance) of 0.001. The iid2d-model fixes random effect priors to a two-dimensional Normal Wishart distribution. For recruitment selection analyses, we controlled for overdispersion by adding an observation-level random effect with log-gamma prior with shape ($\alpha = 0.5$) and mean value (β = variance (offspring recruitment) * α).

We interpreted estimates of fixed effects as statistically significant if their 95% CIs did not overlap zero. Statistical significance of average selection was inferred from models fitting absolute fitness as the response variable, as those fully fulfilled distributional assumptions (SI Appendix Table S1), while standardized selection gradients are instead provided in the main text (Table 1). Because variance estimates are always zero-positive, the statistical significance of random slope variance (indicative of variation in selection) was instead calculated by generating a null distribution for the amount of variance expected by chance. We calculated this null distribution for each hierarchical level separately (i.e., population, plot, year, population-year or plot-year) by permuting the focal levels (e.g., 188 plot-years) associated with a focal variance component (e.g., plot × year), and rerunning each analysis 1000 times (64). We subsequently calculated the proportion of 1000 null values that were greater than the observed variance as a value of P. Values of $P < 0.05$ were considered statistically significant.

Data availability. Data and code to reproduce statistical analyses and Fig. 2 are available on Dryad repository: https://doi.org/10.5061/dryad.mkkwh70z8 (65)

ACKNOWLEDGMENTS

We thank the editor and two anonymous referees for constructive feedback. We also thank all the field assistants, students and members of the different research groups for help in data collection, with particular thanks to Frank Adriaensen for management of the Antwerp dataset. We are grateful to Håvard Rue and Yimen Araya-Ajoy for statistical advice. All data collected were ethically approved by the Regierung von Oberbayern (ROB-55.2-2532.Vet_02-17-215) in Bavaria, by the Dutch legal entity Dier Experimenten Comissie (DEC) and Central Authority for Scientific Procedures on Animals (CCD) (NIOO 14.11 and AVD-801002017831) in Westerheide, by the Animal Experimental Committee of the University of Groningen (DEC-

4114 B) in Lauwersmeer, by the Flemish Ministry (ANB/BL/FF/10-00791, V13-00214) in Antwerp and by the British Trust for Ornithology and Natural England licenses (20093577, 20104175, 20114175, 20131205) in Wytham Woods. A. Mouchet and N.J. Dingemanse were funded by the German Science Foundation (grant no. DI 1694/1-1) and A. Mouchet was a member of the International Max Planck Research School for Organismal Biology. E. Cole was supported by a BBSRC fellowship. M. Nicolaus was funded by the Netherlands Organization for Scientific Research (grant no. ALWOP.2014.109). E. Matthysen and T. Van Overveld were funded by grants from FWO and UA-BOF. J. Quinn was supported by a Royal Society research grant and an ERC Consolidator Grant (project no. 617509). A. M. Roth was funded by scholarships from St. Catherine's College, University of Oxford. K. van Oers was funded by the Netherlands Organization for Scientific Research (grant no. ALWOP.314).

REFERENCES

1. K. van Oers, G. De Jong, A. J. Van Noordwijk, B. Kempenaers, P. J. Drent, Contribution of Genetics to the Study of Animal Personalities: A Review of Case Studies. Behaviour 142, 1185–1206 (2005).
2. N. J. Dingemanse, M. Wolf, Recent models for adaptive personality differences: A review. Philos. Trans. R. Soc. B Biol. Sci. 365, 3947–3958 (2010).
3. M. Wolf, F. J. Weissing, An explanatory framework for adaptive personality differences. Philos. Trans. R. Soc. B Biol. Sci. 365, 3959–3968 (2010).
4. M. Wolf, G. S. van Doorn, F. J. Weissing, Evolutionary emergence of responsive and unresponsive personalities. Proc. Natl. Acad. Sci. 105, 15825–15830 (2008).
5. A. M. Roth, et al., Sexual selection and personality: Individual and group-level effects on mating behaviour in red junglefowl. J. Anim. Ecol., 1–19 (2021).
6. M. Wolf, J. M. McNamara, On the Evolution of Personalities via Frequency-Dependent Selection. Am. Nat. 179, 679–692 (2012).
7. A. M. Siepielski, J. D. Dibattista, S. M. Carlson, It's about time: The temporal dynamics of phenotypic selection in the wild. Ecol. Lett. 12, 1261–1276 (2009).
8. A. M. Siepielski, et al., The spatial patterns of directional phenotypic selection. Ecol. Lett. 16, 1382–1392 (2013).
9. S. A. Foster, J. A. Endler, Geographic variation in behavior. Perspectives on evolutionary mechanisms., S. A. Foster, J. A. Endler, Eds. (Oxford University Press, 1999).
10. P. Edelaar, D. I. Bolnick, Non-random gene flow: An underappreciated force in evolution and ecology. Trends Ecol. Evol. 27, 659–665 (2012).
11. J. Cote, J. Clobert, T. Brodin, S. Fogarty, A. Sih, Personality-dependent dispersal: Characterization, ontogeny and consequences for spatially structured populations. Philos. Trans. R. Soc. B Biol. Sci. 365, 4065–4076 (2010).
12. N. J. Dingemanse, C. Both, P. J. Drent, K. Van Oers, A. J. Van Noordwijk, Repeatability and heritability of exploratory behaviour in great tits from the wild. Anim. Behav. 64, 929–938 (2002).
13. N. J. Dingemanse, et al., Variation in personality and behavioural plasticity across four populations of the great tit Parus major. J. Anim. Ecol. 81, 116–126 (2012).
14. P. Korsten, T. Van Overveld, F. Adriaensen, E. Matthysen, Genetic integration of local dispersal and exploratory behaviour in a wild bird. Nat. Commun. 4, 1–7 (2013).

15. M. Nicolaus, R. Piault, R. Ubels, J. M. Tinbergen, N. J. Dingemanse, The correlation between coloration and exploration behaviour varies across hierarchical levels in a wild passerine bird. J. Evol. Biol. 29, 1780–1792 (2016).
16. A. W. Santure, et al., Replicated analysis of the genetic architecture of quantitative traits in two wild great tit populations. Mol. Ecol. 24, 6148–6162 (2015).
17. N. J. Dingemanse, C. Both, A. J. Van Noordwijk, A. L. Rutten, P. J. Drent, Natal dispersal and personalities in great tits (Parus major). Proc. R. Soc. B Biol. Sci. 270, 741–747 (2003).
18. J. L. Quinn, S. C. Patrick, S. Bouwhuis, T. A. Wilkin, B. C. Sheldon, Heterogeneous selection on a heritable temperament trait in a variable environment. J. Anim. Ecol. 78, 1203–1215 (2009).
19. P. Edelaar, A. M. Siepielski, J. Clobert, Matching habitat choice causes directed gene flow: A neglected dimension in evolution and ecology. Evolution (N. Y). 62, 2462–2472 (2008).
20. B. Holtmann, E. S. A. Santos, C. E. Lara, S. Nakagawa, Personality-matching habitat choice, rather than behavioural plasticity, is a likely driver of a phenotype–environment covariance. Proc. R. Soc. B Biol. Sci. 284, 1–9 (2017).
21. M. Lemoine, et al., Low but contrasting neutral genetic differentiation shaped by winter temperature in European great tits. Biol. J. Linn. Soc. 118, 668–685 (2016).
22. A. Duckworth, A. V Badyaev, Coupling of dispersal and aggression facilitates the rapid range expansion of a passrine bird. Proc. Natl. Acad. Sci. 104, 15017–15022 (2007).
23. M. E. M. Verbeek, P. De Goede, P. J. Drent, P. R. Wiepkema, Individual behavioural characteristics and dominance in aviary groups of great tits. Behaviour 136, 23–48 (1999).
24. J. Wright, G. H. Bolstad, Y. G. Araya-Ajoy, N. J. Dingemanse, Life-history evolution under fluctuating density-dependent selection and the adaptive alignment of pace-of-life syndromes. Biol. Rev. 94, 230–247 (2019).
25. A. A. Munson, C. Jones, H. Schraft, A. Sih, You're Just My Type: Mate Choice and Behavioral Types. Trends Ecol. Evol. 35, 823–833 (2020).
26. M. Nicolaus, P. Edelaar, Comparing the consequences of natural selection, adaptive phenotypic plasticity, and matching habitat choice for phenotype–environment matching, population genetic structure, and reproductive isolation in meta-populations. Ecol. Evol. 8, 3815–3827 (2018).
27. S. Verhulst, C. M. Perrins, R. Riddington, Natal dispersal of Great Tits in a patchy environment. Ecology 78, 864–872 (1997).

28. M. Nicolaus, J. M. Tinbergen, R. Ubels, C. Both, N. J. Dingemanse, Density fluctuations represent a key process maintaining personality variation in a wild passerine bird. Ecol. Lett. 19, 478–486 (2016).
29. B. E. Sæther, et al., The extended Moran effect and large-scale synchronous fluctuations in the size of great tit and blue tit populations. J. Anim. Ecol. 76, 315–325 (2007).
30. A. C. Perdeck, M. E. Visser, J. H. van Balen, Great tit Parus major survival and the beech-crop cycle. Ardea 88 (1), 99–108 (2000).
31. G. Vacchiano, et al., Spatial patterns and broad-scale weather cues of beech mast seeding in Europe. New Phytol. 215, 595–608 (2017).
32. D. J. Weese, S. P. Gordon, A. P. Hendry, M. T. Kinnison, Spatiotemporal variation in linear natural selection on body color in wild guppies (poecilia reticulata). Evolution (N. Y). 64, 1802–1815 (2010).
33. J. L. Quinn, E. F. Cole, J. Bates, R. W. Payne, W. Cresswell, Personality predicts individual responsiveness to the risks of starvation and predation. Proc. R. Soc. B Biol. Sci. 279, 1919–1926 (2012).
34. R. N. Abbey-Lee, N. J. Dingemanse, Adaptive individual variation in phenological responses to perceived predation levels. Nat. Commun. 10 (2019).
35. R. E. Rollins, et al., Repeatable differences in exploratory behaviour predict tick infestation probability in wild great tits. Behav. Ecol. Sociobiol. 75 (2021).
36. M. Wolf, G. S. Van Doorn, O. Leimar, F. J. Weissing, Life-history trade-offs favour the evolution of animal personalities. Nature 447, 581–584 (2007).
37. D. Réale, et al., Personality and the emergence of the pace-of-life syndrome concept at the population level. Philos. Trans. R. Soc. B Biol. Sci. 365, 4051–4063 (2010).
38. M. Nicolaus, et al., Does coping style predict optimization? An experimental test in a wild passerine bird. Proc. R. Soc. B Biol. Sci. 282 (2015).
39. Y. G. Araya-Ajoy, et al., Sources of (co)variation in alternative siring routes available to male great tits (Parus major). Evolution (N. Y). 70, 2308–2321 (2016).
40. N. J. Dingemanse, M. Moiron, Y. G. Araya-Ajoy, A. Mouchet, R. N. Abbey-Lee, Individual variation in age-dependent reproduction: fast explorers live fast but senesce young? J. Anim. Ecol. 88, 1–13 (2019).
41. R. Royauté, M. A. Berdal, C. R. Garrison, N. A. Dochtermann, Paceless life? A meta-analysis of the pace-of-life syndrome hypothesis. Behav. Ecol. Sociobiol. 72, 1–10 (2018).
42. M. Moiron, K. L. Laskowski, P. T. Niemelä, Individual differences in behaviour explain variation in survival: a meta-analysis. Ecol. Lett. 23, 399–408 (2020).

43. P. O. Montiglio, M. Dammhahn, G. Dubuc Messier, D. Réale, The pace-of-life syndrome revisited: the role of ecological conditions and natural history on the slow-fast continuum. Behav. Ecol. Sociobiol. 72 (2018).
44. K. L. Laskowski, M. Moiron, P. Niemelä, Integrating Behavior in Life-History Theory: Allocation versus Acquisition? Trends Ecol. Evol., 1–7 (2020).
45. N. J. Dingemanse, C. Both, P. J. Drent, J. M. Tinbergen, Fitness consequences of avian personalities in a fluctuating environment. Proc. R. Soc. B Biol. Sci. 271, 847–852 (2004).
46. P. H. Harvey, P. J. Greenwood, C. M. Perrins, Breeding area fidelity of great tits (Parus major). J. Anim. Ecol. 48, 305–313 (1979).
47. J. M. Tinbergen, Biased estimates of fitness consequences of brood size manipulation through correlated effects on natal dispersal. J. Anim. Ecol. 74, 1112–1120 (2005).
48. A. Culina, Connecting the data landscape of long-term ecological studies: the SPI-Birds data hub. J. Anim. Ecol. (2020) https:/doi.org/10.1111/1365-2656.13388.
49. E. Matthysen, Boundary effects on dispersal between habitat patches by forest birds (Parus major, P. caeruleus). Landsc. Ecol. 17, 509–515 (2002).
50. P. J. Drent, K. Van Oers, A. J. Van Noordwijk, Realized heritability of personalities in the great tit (Parus major). Proc. R. Soc. B Biol. Sci. 270, 45–51 (2003).
51. R. N. Abbey-Lee, K. J. Mathot, N. J. Dingemanse, Behavioral and morphological responses to perceived predation risk: A field experiment in passerines. Behav. Ecol. 27, 857–864 (2016).
52. M. E. Visser, C. M. Lessells, The costs of egg production and incubation in great tits (Parus major). Proc. R. Soc. B Biol. Sci. 268, 1271–1277 (2001).
53. A. Qvarnström, J. E. Brommer, L. Gustafsson, Testing the genetics underlying the co-evolution of mate choice and ornament in the wild. Nature 441, 84–86 (2006).
54. P. T. Niemelä, N. J. Dingemanse, On the usage of single measurements in behavioural ecology research on individual differences. Anim. Behav. 145, 99–105 (2018).
55. N. J. Dingemanse, N. A. Dochtermann, Quantifying individual variation in behaviour: Mixed-effect modelling approaches. J. Anim. Ecol. 82, 39–54 (2013).
56. J. G. A. Martin, D. Réale, Temperament, risk assessment and habituation to novelty in eastern chipmunks, Tamias striatus. Anim. Behav. 75, 309–318 (2008).
57. R. Lande, S. Arnold, The Measurement of Selection on Correlated Characters. Evolution (N. Y). 37, 1210–1226 (1983).
58. C. Both, M. E. Visser, N. Verboven, Density-dependent recruitment rates in great tits: The importance of being heavier. Proc. R. Soc. B Biol. Sci. 266, 465–469 (1999).

59. C. Both, J. Tinbergen, M. Visser, Adaptive Density Dependence of Avian Clutch Size. Ecology 81, 3391–3403 (2000).
60. S. P. De Lisle, E. I. Svensson, On the standardization of fitness and traits in comparative studies of phenotypic selection. Evolution (N. Y). 71, 2313–2326 (2017).
61. J. R. Stinchcombe, A. F. Agrawal, P. A. Hohenlohe, S. J. Arnold, M. W. Blows, Estimating nonlinear selection gradients using quadratic regression coefficients: Double or nothing? Evolution (N. Y). 62, 2435–2440 (2008).
62. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
63. H. Rue, S. Martino, N. Chopin, Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B Stat. Methodol. 71, 319–392 (2009).
64. P. I. Good, Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses. (2000).
65. A. Mouchet et al. (2021). Data and code for Heterogeneous selection on exploration behavior within and among West European populations of a passerine bird. Dryad. https://doi.org/10.5061/dryad.mkkwh70z8

SUPPORTING INFORMATION FOR CHAPTER 2

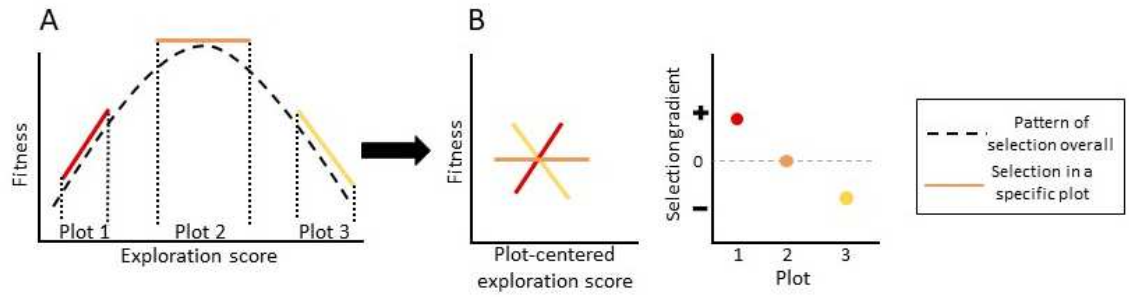


Fig. S1. A hypothetical scenario illustrating how failure to model nonlinear selection can result in the appearance of heterogeneous selection where none exists. A) Individuals in Plot 1 are relatively slower than individuals in Plot 2, which are themselves relatively slower than individuals in Plot 3. B) Failure to acknowledge this nonlinear selection would consequently result in random regression analyses estimating significant random slopes across plots when exploration scores are within-plot centered, leading to the false conclusion that there is heterogeneous selection. This artefact can be avoided by fitting the linear and quadratic components of each plot's mean exploration behavior as fixed effects into the model. Applied to the illustrated scenario, such a model would demonstrate evidence for stabilizing selection but no evidence for heterogeneous selection among plots, because none of the plots differ in selection when this form of nonlinearity is statistically controlled for. We expanded upon this idea and accounted for nonlinear selection at each hierarchical level (population, plot, year, population-year and plot-year) by fitting linear and quadratic components of each level's mean exploration behavior as fixed effects in Table S3. This expanded model did not exhibit a decreased variance in random slopes, implying that our estimates of heterogeneous selection printed in the main text (Table 2) did not stem from an artefact from failing to model level-specific patterns of nonlinearity.

Table S1. Linear and nonlinear selection on exploration behavior estimated using integrative fitness, adult survival or offspring recruitment as the fitness metric. Estimates for exploration behavior are derived from random regression models using unstandardized fitness data, while exploration scores were standardized over the entire dataset. We print variance components (σ^2) estimating variation in selection at various ecological levels, and intercept-slope correlations (ρ). P-values obtained from permutation tests are the probability that variation in selection is greater than expected by chance. All estimates are presented with their 95% credible interval (95% CI).

	Integrative fitness		**Survival**		**Recruitment**	
Fixed effects	Estimate (95% CI)		Estimate (95% CI)		Estimate (95% CI)	
Intercept	0.61 (0.14, 1.09)		-0.52 (-1.11, 0.05)		-0.8 (-1.5, -0.11)	
Exploration						
Linear term (β)	0.02 (-0.44, 0.47)		0.03 (-0.47, 0.54)		0.02 (-0.5, 0.53)	
Quadratic term (γ)	0.00 (-0.01, 0.01)		0.03 (-0.02, 0.07)		-0.06 (-0.1, -0.02)	
Random Effects	σ^2 (95% CI)	p-value	σ^2 (95% CI)	p-value	σ^2 (95% CI)	p-value
Exploration × Population						
σ^2 intercept	0.18 (0.06, 0.45)	<0.001	0.22 (0.07, 0.59)	<0.001	0.32 (0.09, 0.87)	<0.001
σ^2 slope	0.17 (0.06, 0.44)	<0.001	0.19 (0.07, 0.48)	<0.001	0.19 (0.07, 0.49)	<0.001
ρ intercept × slope	0.04 (-0.49, 0.56)		-0.03 (-0.56, 0.50)		0.02 (-0.54, 0.58)	
Exploration × Plot						
σ^2 intercept	0.05 (0.03, 0.09)	<0.001	0.12 (0.06, 0.22)	0.003	0.14 (0.07, 0.27)	<0.001
σ^2 slope	0.04 (0.03, 0.07)	<0.001	0.08 (0.04, 0.14)	0.72	0.07 (0.04, 0.13)	0.115
ρ intercept × slope	-0.02 (-0.36, 0.33)		-0.06 (-0.48, 0.37)		-0.09 (-0.49, 0.32)	
Exploration × Year						
σ^2 intercept	0.08 (0.04, 0.15)	<0.001	0.15 (0.07, 0.31)	<0.001	0.21 (0.08, 0.47)	<0.001
σ^2 slope	0.07 (0.03, 0.12)	<0.001	0.08 (0.04, 0.16)	0.23	0.08 (0.04, 0.16)	<0.001
ρ intercept × slope	-0.01 (-0.41, 0.41)		0.004 (-0.44, 0.43)		0.06 (-0.42, 0.54)	
Exploration × Population × Year						
σ^2 intercept	0.05 (0.03, 0.08)	<0.001	0.13 (0.07, 0.24)	<0.001	0.19 (0.1, 0.34)	<0.001
σ^2 slope	0.04 (0.02, 0.06)	<0.001	0.07 (0.04, 0.11)	0.13	0.06 (0.03, 0.1)	<0.001
ρ intercept × slope	-0.003 (-0.31, 0.31)		-0.02 (-0.41, 0.38)		0.06 (-0.35, 0.45)	
Exploration × Plot × Year						
σ^2 intercept	0.04 (0.03, 0.06)	<0.001	0.13 (0.06, 0.24)	0.007	0.14 (0.08, 0.22)	<0.001
σ^2 slope	0.03 (0.02, 0.04)	0.03	0.09 (0.05, 0.15)	0.43	0.06 (0.03, 0.09)	0.09
ρ intercept × slope	0.05 (-0.19, 0.31)		0.09 (-0.34, 0.50)		0.03 (-0.34, 0.39)	
Residual	0.38 (0.36, 0.39)		-		-	
Overdispersion	-		-		0.22 (0.15, 0.31)	

Table S2. Exploration behavior of yearling immigrants versus yearling local recruits. Birds were categorized as immigrants if they had not been ringed as nestlings in the study area (regardless of plot identity). Only yearlings were included in this analysis to avoid any biases due to changes in exploration behavior with age (1). Estimates are derived from a random intercept regression model fitted with Gaussian error distribution. Exploration score was corrected for seasonal change in behavior from July to March, considering July 1st as the reference date (2; see also main text). We print estimates for fixed effects and variance components, with their 95% credible interval (95% CI). Local recruits are the reference category.

Fixed effects	β (95% CI)
Intercept	-0.08 (-0.28, 0.12)
Immigrant	0.18 (0.06, 0.30)
Random Effects	σ^2 (95% CI)
Population	0.02 (0.00, 0.14)
Plot	0.003 (0.005, 0.02)
Year	0.01 (0.00, 0.07)
Population × Year	0.05 (0.02, 0.12)
Plot × Year	0.01 (0.007, 0.03)
Residual	0.96 (0.91, 1.02)

Table S3. Linear and nonlinear selection on exploration behavior estimated using integrative fitness as the combination of adult survival and half of the annual number of fledglings or annual number of fledglings as the fitness metric. Estimates for exploration behavior are derived from random regression models using unstandardized fitness data, while exploration scores were standardized over the entire dataset. We print variance components (σ^2) estimating variation in selection at various ecological levels, and intercept-slope correlations (ρ). All estimates are presented with their 95% credible interval (95% CI).

	Integrative fitness	Number of fledglings
Fixed effects	β (95% CI)	β (95% CI)
Intercept	3.39 (2.75, 4.01)	1.68 (1.20, 2.16)
Exploration		
Linear term (β)	0.02 (-0.47, 0.51)	0.01 (-0.44, 0.46)
Quadratic term (γ)	-0.01 (-0.04, 0.03)	-0.01 (-0.02, 0.00)
Random Effects	σ^2 (95% CI)	σ^2 (95% CI)
Exploration x Population		
σ^2 intercept	0.25 (0.08, 0.70)	0.19 (0.07, 0.47)
σ^2 slope	0.18 (0.06, 0.47)	0.18 (0.06, 0.45)
ρ intercept x slope	0.1 (-0.44, 0.61)	0.02 (-0.48, 0.51)
Exploration x Plot		
σ^2 intercept	0.18 (0.09, 0.32)	0.06 (0.03, 0.10)
σ^2 slope	0.07 (0.04, 0.13)	0.04 (0.02, 0.07)
ρ intercept x slope	0.17 (-0.26, 0.55)	0.10 (-0.23, 0.42)
Exploration x Year		
σ^2 intercept	0.26 (0.11, 0.57)	0.09 (0.05, 0.17)
σ^2 slope	0.08 (0.04, 0.15)	0.06 (0.03, 0.11)
ρ intercept x slope	-0.04 (-0.56, 0.43)	-0.05 (-0.45, 0.36)
Exploration x Pop-Year		
σ^2 intercept	0.20 (0.09, 0.34)	0.05 (0.03, 0.09)
σ^2 slope	0.06 (0.03, 0.09)	0.03 (0.02, 0.05)
ρ intercept x slope	-0.09 (-0.53, 0.31)	-0.05 (-0.34, 0.25)
Exploration x Plot-Year		
σ^2 intercept	0.17 (0.10, 0.27)	0.04 (0.03, 0.06)
σ^2 slope	0.06 (0.04, 0.10)	0.02 (0.02, 0.03)
ρ intercept x slope	0.08 (-0.30, 0.42)	0.04 (-0.21, 0.27)
Residual	2.41 (2.32, 2.52)	
Overdispersion		0.09 (0.08, 0.10)
Ecological level	R (95% CI)	R (95% CI)
Population	0.38 (0.30,0.50)	0.50 (0.40,0.63)
Plot	0.17 (0.14,0.19)	0.13 (0.09,0.15)
Year	0.18 (0.16,0.19)	0.19 (0.16,0.20)
Population × Year	0.13 (0.10,0.16)	0.11 (0.08,0.13)
Plot × Year	0.14 (0.11,0.17)	0.07 (0.05,0.11)

Table S4. Linear and nonlinear selection on exploration behavior estimated at each hierarchical level, for each of three alternative fitness metrics. Estimates were derived from random regression models using exploration scores standardized over the entire dataset and best linear unbiased predictors (BLUPs) for exploration behavior at each hierarchical level. We print variance components (σ^2) estimating variation in selection at different hierarchical levels, and associated intercept-slope correlations (ρ). All estimates are printed with 95% credible interval (95% CI). BLUPs account for possible variation in phenotypes among hierarchical levels that could result in the detection of variation in selection where none exists (illustrated in Fig.S1). The analyses printed here had random effect estimates similar to those shown in Table S1, indicating that variation in selection was not an artefact of non-random settlement of phenotypes. We therefore reject this explanation.

	Integrative fitness	**Survival**	**Recruitment**
Fixed Effects	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)
Intercept	0.60 (-0.08, 1.27)	-0.52 (-1.30, 0.24)	-1.13 (-1.99, -0.19)
Exploration			
Linear term (β)	0.02 (-0.46, 0.49)	0.01 (-0.05, 0.07)	0.05 (-0.47, 0.57)
BLUP for population	0.01 (-0.65, 0.68)	1.21 (-5.53, 7.97)	-0.17 (-1.46, 1.07)
BLUP for plot	-0.11 (-8.48, 8.25)	-1.17 (-3.49, 1.13)	-6.03 (-26, 13.84)
BLUP for year	-0.06 (-1.39, 1.27)	-0.11 (-1.40, 1.20)	1.06 (-1.56, 3.7)
BLUP for population × year	-0.24 (-6.05, 5.54)	0.57 (-3.35, 4.48)	1.9 (-5.97, 9.53)
BLUP for plot-year	0.0004 (-0.01, 0.01)	1.25 (-48.79, 51.22)	-0.06 (-0.10, -0.03)
Quadratic term (γ)	-0.07 (-3.56, 3.43)	0.04 (-0.01, 0.08)	-1.07 (-6.82, 4.63)
BLUP² for population	3.54 (-36.47, 42.96)	-3.15 (-48.16, 42.10)	32.16 (-18.71, 79.58)
BLUP² for plot	2.35 (-11.03, 15.73)	20.57 (-1.98, 43.14)	-5.79 (-30.38, 18.84)
BLUP² for year	0.16 (-4.01, 4.31)	3.46 (-5.85, 12.74)	-2.11 (-9.69, 5.51)
BLUP² for population × year	-10.47 (-49.44, 28.98)	-11.29 (-43.77, 21.53)	-16.85 (-65.27, 32.33)
BLUP² for plot-year	-2.38 (-64.35, 59.54)	0.07 (-62.02, 62.10)	-0.69 (-62.71, 61.27)
Random Effects	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)
Exploration × Population			
σ^2 intercept	0.23 (0.07, 0.65)	0.25 (0.07, 0.70)	0.36 (0.09, 1.12)
σ^2 slope	0.18 (0.06, 0.45)	0.43 (0.09, 1.46)	0.19 (0.06, 0.48)
ρ intercept × slope	-0.02 (-0.56, 0.53)	-0.02 (-0.61, 0.57)	-0.02 (-0.59, 0.55)
Exploration × Plot			
σ^2 intercept	0.06 (0.03, 0.09)	0.11 (0.05, 0.21)	0.19 (0.09, 0.37)
σ^2 slope	0.04 (0.03, 0.07)	0.41 (0.09, 1.34)	0.09 (0.04, 0.16)
ρ intercept × slope	-0.03 (-0.36, 0.32)	0.04 (-0.51, 0.56)	-0.25 (-0.63, 0.18)
Exploration × Year			
σ^2 intercept	0.09 (0.04, 0.16)	0.15 (0.06, 0.31)	0.22 (0.08, 0.48)
σ^2 slope	0.06 (0.03, 0.12)	0.40 (0.09, 1.30)	0.08 (0.04, 0.15)
ρ intercept × slope	0.00 (-0.4, 0.4)	0.08 (-0.49, 0.64)	0.09 (-0.39, 0.56)
Exploration × Population × Year			
σ^2 intercept	0.05 (0.03, 0.08)	0.13 (0.07, 0.25)	0.16 (0.08, 0.3)
σ^2 slope	0.04 (0.02, 0.06)	0.39 (0.09, 1.21)	0.05 (0.03, 0.09)
ρ intercept × slope	0.01 (-0.32, 0.32)	-0.08 (-0.62, 0.48)	0.07 (-0.33, 0.45)
Exploration × Plot × Year			
σ^2 intercept	0.04 (0.03, 0.06)	0.12 (0.06, 0.22)	0.2 (0.12, 0.31)
σ^2 slope	0.03 (0.02, 0.04)	0.42 (0.09, 1.45)	0.06 (0.04, 0.1)
ρ intercept × slope	0.05 (-0.19, 0.31)	-0.01 (-0.55, 0.53)	-0.04 (-0.4, 0.35)
Residual	0.38 (0.36, 0.39)		
Overdispersion			0.001 (0.00, 0.004)

Table S5. Estimates of linear and nonlinear selection on exploration behavior with exploration scores standardized within each unique combination of plot and year (plot-year) for three alternative fitness metrics. We print variance components (σ^2) estimating variation in selection at different hierarchical levels, and associated intercept-slope correlations (ρ). All estimates are printed with 95% credible interval (95% CI).

	Integrative fitness	**Survival**	**Recruitment**
Fixed effects	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)
Intercept	0.61 (0.12, 1.10)	-0.53 (-1.09, 0.02)	-1.27 (-1.97, -0.57)
Exploration			
Linear term (β)	0.01 (-0.46, 0.49)	0.04 (-048, 0.56)	0.03 (-0.48, 0.54)
Quadratic term (γ)	0.00 (-0.01, 0.02)	0.04 (-0.01, 0.09)	-0.06 (-0.10, -0.02)
Random effects	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)
Exploration × Population			
σ^2 intercept	0.19 (0.06, 0.47)	0.21 (0.07, 0.53)	0.31 (0.09, 0.85)
σ^2 slope	0.18 (0.06, 0.46)	0.2 (0.07, 0.50)	0.20 (0.07, 0.49)
ρ intercept × slope	0.01 (-0.53, 0.54)	-0.03 (-0.56, 0.5)	0.03 (-0.54, 0.59)
Exploration × Plot			
σ^2 intercept	0.05 (0.03, 0.08)	0.12 (0.06, 0.23)	0.17 (0.08, 0.31)
σ^2 slope	0.04 (0.03, 0.07)	0.09 (0.04, 0.15)	0.08 (0.04, 0.15)
ρ intercept × slope	-0.02 (-0.36, 0.31)	-0.12 (-0.54, 0.32)	-0.17 (-0.57, 0.27)
Exploration × Year			
σ^2 intercept	0.08 (0.04, 0.15)	0.15 (0.07, 0.30)	0.23 (0.09, 0.50)
σ^2 slope	0.07 (0.03, 0.12)	0.09 (0.04, 0.16)	0.08 (0.04, 0.16)
ρ intercept × slope	0.00 (-0.41, 0.40)	0.01 (-0.43, 0.44)	0.08 (-0.39, 0.54)
Exploration × Population × Year			
σ^2 intercept	0.05 (0.03, 0.08)	0.13 (0.07, 0.24)	0.20 (0.10, 0.36)
σ^2 slope	0.04 (0.02, 0.06)	0.07 (0.04, 0.12)	0.06 (0.03, 0.10)
ρ intercept × slope	-0.01 (-0.31, 0.29)	-0.04 (-0.43, 0.36)	0.08 (-0.39, 0.54)
Exploration × Plot-Year			
σ^2 intercept	0.04 (0.03, 0.05)	0.12 (0.06, 0.22)	0.17 (0.09, 0.27)
σ^2 slope	0.03 (0.02, 0.04)	0.09 (0.05, 0.17)	0.06 (0.04, 0.10)
ρ intercept × slope	0.02 (-0.24, 0.27)	0.07 (-0.36, 0.49)	-0.04 (-0.42, 0.35)
Residual	0.38 (0.36, 0.39)		
Overdispersion			0.30 (0.22, 0.38)

Table S6. Linear and nonlinear density-dependent selection on exploration behavior, where either adult survival, offspring recruitment, or integrative fitness, was used as fitness metric. Estimates for selection on exploration behavior were derived from random regression models using unstandardized fitness data, while exploration scores were standardized over the entire dataset. We print variance components (σ^2) estimating variation in selection at different hierarchical levels with their associated intercept-slope correlations (ρ). All estimates are printed with their 95% credible interval (95% CI). We calculated breeding density for each combination of plot and year (plot-year) as the number of breeding pairs per hectare (pairs/ha). Each plot area was calculated by manual delimitation, following the outer nest box line and excluding open fields, using the software QGIS (v. 2.18.25) (3). We expressed breeding density of a focal plot as a deviation of the focal population's mean density (within-population density); we also expressed the population's mean density as a deviation of the population-mean value estimated across the five populations (among-population density). This enabled us to estimate whether selection on exploration behavior varied within populations as a function of micro-spatiotemporal variation in density and among populations as a function of macro-spatial variation in density. We performed this analysis to test whether variation in breeding density underpinned the variance in selection estimated in Table S1. This would be the case if models controlling for personality-related density-dependent selection would exhibit decreased random slope variance (compared to Table S1), which was not the case for any hierarchical level.

	Integrative fitness	**Survival**	**Recruitment**
Fixed effects	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)
Intercept	0.59 (0.03, 1.16)	-0.57 (-1.22, 0.08)	-1.29 (-2.00, -0.58)
Exploration			
Linear term (β)	0.01 (-0.51, 0.53)	0.04 (-0.54, 0.63)	0.03 (-0.52, 0.57)
Quadratic term (γ)	0.00 (-0.01, 0.01)	0.04 (-0.01, 0.09)	-0.06 (-0.10, -0.02)
Within-population density	-0.13 (-0.28, 0.02)	-0.06 (-0.40, 0.27)	-0.34 (-0.59, -0.10)
Among-population density	0.02 (-0.33, 0.37)	0.03 (-0.35, 0.42)	0.03 (-0.30, 0.37)
Within- x among-population density	0.03 (-0.02, 0.08)	0.00 (-0.12, 0.11)	0.09 (0.01, 0.16)
Exploration x within-population density	-0.05 (-0.18, 0.08)	-0.11 (-0.38, 0.17)	-0.02 (-0.20, 0.16)
Exploration x among-population density	-0.01 (-0.33, 0.32)	-0.01 (-0.36, 0.35)	0.00 (-0.27, 0.27)
Exploration² x within-population density	-0.01 (-0.05, 0.04)	-0.06 (-0.21, 0.08)	0.02 (-0.07, 0.11)
Exploration² x among-population density	0.00 (-0.01, 0.01)	0.03 (0.00, 0.06)	0.00 (-0.03, 0.02)
Exploration x within- x among-population density	0.02 (-0.03, 0.07)	0.05 (-0.05, 0.15)	0.01 (-0.04, 0.06)
Exploration² x within- x among-population density	0.00 (-0.01, 0.02)	0.04 (-0.01, 0.09)	-0.01 (-0.03, 0.02)
Random effects	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)
Exploration x Population			
σ^2 intercept	0.22 (0.07, 0.58)	0.26 (0.08, 0.71)	0.34 (0.09, 0.94)
σ^2 slope	0.21 (0.07, 0.57)	0.23 (0.07, 0.61)	0.22 (0.07, 0.58)
ρ intercept x slope	0.04 (-0.52, 0.59)	0.00 (-0.57, 0.56)	0.03 (-0.55, 0.61)
Exploration x Plot			
σ^2 intercept	0.05 (0.03, 0.09)	0.14 (0.06, 0.26)	0.16 (0.07, 0.30)
σ^2 slope	0.05 (0.03, 0.07)	0.09 (0.04, 0.16)	0.09 (0.05, 0.17)
ρ intercept x slope	-0.04 (-0.37, 0.30)	-0.11 (-0.54, 0.35)	-0.23 (-0.62, 0.21)
Exploration x Year			
σ^2 intercept	0.08 (0.04, 0.14)	0.15 (0.06, 0.3)	0.21 (0.08, 0.46)
σ^2 slope	0.07 (0.03, 0.12)	0.09 (0.04, 0.16)	0.08 (0.04, 0.16)
ρ intercept x slope	-0.01 (-0.42, 0.39)	-0.03 (-0.48, 0.42)	0.08 (-0.40, 0.54)
Exploration x Population x Year			
σ^2 intercept	0.05 (0.03, 0.07)	0.13 (0.07, 0.23)	0.20 (0.10, 0.35)
σ^2 slope	0.04 (0.02, 0.06)	0.07 (0.04, 0.12)	0.06 (0.03, 0.10)
ρ intercept x slope	0.00 (-0.33, 0.32)	-0.07 (-0.46, 0.33)	0.07 (-0.35, 0.47)
Exploration x Plot x Year			
σ^2 intercept	0.04 (0.03, 0.06)	0.13 (0.06, 0.23)	0.16 (0.09, 0.27)
σ^2 slope	0.03 (0.02, 0.04)	0.10 (0.05, 0.17)	0.06 (0.04, 0.10)
ρ intercept x slope	0.06 (-0.20, 0.31)	0.07 (-0.37, 0.48)	-0.06 (-0.44, 0.33)
Residual	0.38 (0.36, 0.39)		
Overdispersion			0.30 (0.22, 0.39)

SI REFERENCES

1. N. J. Dingemanse, M. Moiron, Y. G. Araya-Ajoy, A. Mouchet, R. N. Abbey-Lee, Individual variation in age-dependent reproduction: fast explorers live fast but senesce young? J. Anim. Ecol. 88, 1–13 (2019).
2. N. J. Dingemanse, C. Both, P. J. Drent, K. Van Oers, A. J. Van Noordwijk, Repeatability and heritability of exploratory behaviour in great tits from the wild. Anim. Behav. 64, 929–938 (2002).
3. QGIS Development Team (2018). QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org.

Chapter 3

Supplementary food does not drive heterogeneous selection on activity but makes individuals more repeatable

Unpublished manuscript

Supplementary food does not drive heterogeneous selection on activity but make individuals more repeatable

Alexia Mouchet, Niels J. Dingemanse

Abstract

Repeatable among-individual variation in behaviour (called animal personality) may be maintained by heterogeneous selection. Heterogeneous selection is expected to be driven by various ecological factors varying spatiotemporally, such as population density and food availability. Food availability, by positively affecting survival and reproduction, may drive heterogeneous density-dependent selection. Increased competition, through limiting food resources or high population density, may push individuals to behave more differently. Social niche specialisation in highly competitive contexts may result in greater among- and lower within-individual variation in behaviour compared to relaxed competitive contexts. We tested whether selection on behaviour and the amount of among- and within-individual variance varied with food availability in great tits (Parus major). We provided supplementary food during four consecutive years to experimentally manipulate food availability. Unexpectedly, selection on activity did not vary with food availability. Opposite effects of food availability and population density on competition levels may cause heterogeneous food-dependent selection to be counteracted by heterogeneous density-dependent selection. However, among-individual variance in activity was greater, whereas within-individual variance was lower in food-supplemented context. Females also differed in how they changed clutch size with food availability, independent of their activity. Our findings align with the social niche hypothesis and suggest that stronger competition occurred when density was high rather than when food was scarce. Future studies should disentangle the effects of food availability and density to better understand how animal personalities are maintained.

Keywords: animal personality, phenotypic plasticity, timing of reproduction, cross-context correlations, phenotype-environment interactions

INTRODUCTION

Among-individual phenotypic variation has been proposed to be maintained by heterogeneous selection in space and time (Frank & Slatkin 1990; Wolf & Weissing 2010). Spatial variation in selection may favour local adaptation and ultimately population divergence (Felsenstein 1976; Frank & Slatkin 1990; Foster & Endler 1999). Temporal variation in selection instead may counteract local adaptation and favour the coexistence of alternative phenotypes within locations (Haldane & Jayakar 1963; Hedrick 1976; Lande 1976; Byers 2005). The co-occurrence of these two counteracting processes may cancel out the effect of directional selection and explain the persistence of phenotypic variation within most wild populations (Mouchet et al. 2021). Spatial or temporal variation has been well studied and demonstrated on life-history and morphological traits (Siepielski et al. 2009, 2013). However, little is known about heterogeneous selection on behaviour.

Environmental variation may maintain variation in behaviour if the fitness benefits of alternative behavioural tactics differ between environments (Dall et al. 2004; Dingemanse & Wolf 2010; Wolf & Weissing 2010, 2012). This may be the case for behaviours facilitating resource acquisition at the expense of survival, such as aggressiveness, boldness or exploration/activity in novel environments (called risk-taking behaviours). The costs and benefits of risk-taking behaviours are expected to vary with various ecological factors that may alter access to resources, such as climate, predation risk, population density or food availability (Boon et al. 2007; Cote et al. 2008; Le Cœur et al. 2015; Le Galliard et al. 2015; Nicolaus et al. 2016; Abbey-Lee & Dingemanse 2019). For example, in high predation danger environments, risk-taking may increase the probability of depredation. In low food resource environments, risk-taking may be beneficial instead to avoid the risk of starvation. Consequently, within populations, selection on behaviour is expected to vary with micro-spatial and temporal variation in ecological factors, favouring risk-prone individuals in certain conditions, and risk-adverse individuals in other conditions.

Population density and food availability are two factors that vary within most populations both spatially and temporally (Lack 1954). Both factors generally positively correlate: increases in food availability enhance survival probability and reproductive output, and attract immigrants, thereby increasing population density (Perrins 1965; Perdeck et al. 2000; Grøtan et al. 2009). However, both factors affect access to food resources in an opposite manner: an increase in food resources reduces while an increase in density rises competition among individuals. Competitive, risk-prone individuals may be favoured in high densities because of their greater ability to secure resources (Verbeek et al. 1996; Dingemanse & De Goede 2004;

Biro & Stamps 2009). They may be in good reproductive condition earlier than risk-adverse individuals, which may allow them to reproduce earlier and invest more in reproduction. Alternatively, they may be disadvantaged in high densities if costs outweigh benefits of risks due to, for example, higher risk of injuries and parasitism, negative effects of social defeat, or higher metabolic requirements (Verbeek et al. 1999; Wolf et al. 2007; Barber & Dingemanse 2010; Careau & Garland 2012; Niemelä & Dingemanse 2018; Mathot et al. 2019; Rollins et al. 2021). Risk-prone individuals may be favoured in low food availability contexts when aggressive and exploration behaviours may allow securing and finding scarce, clumped resources (Dingemanse & De Goede 2004; Overveld & Matthysen 2010). Again, greater access to food may allow risk-prone individuals to invest more in reproduction. Risk-prone individuals, however, have been shown to rely on social cues in foraging contexts (Groothuis & Carere 2005). This strategy may be more beneficial in high density context, that is, when there are more individuals to copy, and when food is evenly distributed. Risk-prone individuals, thus, may be favoured in high food availability context instead. Alternatively, selection may be weak in good food context such that all phenotypes would be equally favoured. Opposite selection patterns driven by different ecological factors (food availability versus density) on behavioural phenotypes may thus occur, which could explain contrasting findings on the link between fitness and behavioural phenotypes (Smith & Blumstein 2008; Dammhahn et al. 2018; Royauté et al. 2018; Moiron et al. 2020). Studies primarily focused on density-dependence as ecological driver of heterogeneous selection. Given that density correlates with food abundance, density-dependent selection may ultimately be driven by food-dependent selection. A key question is whether food availability affects population density and drives heterogeneous selection on behaviour.

We investigated whether variation in food availability outside the breeding season caused fecundity selection on behaviour to vary in a great tit (Parus major) population. We experimentally manipulated food availability by providing supplementary food in multiple nest box plots over four years. Following a partial cross-over design, all plots were both control and supplemented with food twice over the course of the experiment (Figure 1). This design allowed us to estimate food-dependent selection on behaviour in food-supplemented and non-supplemented contexts. This design additionally allowed us to estimate cross-context among-individual correlations in behavioural and reproductive traits to determine whether individuals all responded similarly to changes in food availability (Figure 1). We used activity in a novel environment assayed in the field as proxy for risk-taking behaviour because this behaviour correlates with aggressiveness in our population (Moiron et al. 2019). Great tit breeding

densities have been demonstrated to increase with winter food availability (e.g., beech mast) (Lack 1964; Balen 1980; Perdeck et al. 2000). We therefore expected that our food manipulation outside the breeding season would increase breeding densities. At the start of our experiment, our predictions on the effects of food availability on selection were based on the results of a study showing selection favouring faster explorer great tits in low densities (Nicolaus et al. 2016). Though behaviour in that study was assayed in a laboratory room, we assumed our field-based behaviour represented the same trait because each have been shown to correlate with other field-based behaviours related to risk-taking (Verbeek et al. 1994; Hollander et al. 2008; Quinn et al. 2009; Amy et al. 2010; Mutzel et al. 2013; Stuber et al. 2013; Moiron et al. 2019). We therefore expected more active individuals to be favoured in non food-supplemented context. We recently showed, however, that laboratory- and field-based behaviour do not correlate in this great tit population (Mouchet & Dingemanse 2021). That is, both assays appeared to elicit the expression of different behaviours that may be under different selection pressures.

Among-individual variation may change not only because of variation in selection pressures, but also because of phenotype-by-environment interactions, whereby phenotypes respond differently to various environments (Gillespie & Turelli 1989; Dingemanse et al. 2010). These interactions may be revealed by variation in plasticity of different phenotypes across environments or by a different amount of individual variance between environments (Dingemanse et al. 2012). Unfavourable conditions are thought to increase genetic and environmental variance due to, respectively, rarer occurrence compared to favourable conditions and greater spatiotemporal variation in the environment (Hoffmann & Merilä 1999). The social niche hypothesis, derived from the niche variation hypothesis, predicts that when competition for resources increases, individuals should specialise on behaviours allowing them to use different resources relative to others in the population (van Valen 1965; Bolnick et al. 2007; Bergmüller & Taborsky 2010). Individual specialisation should consequently decrease overlap in resource use with conspecifics and increase among-individual variation in behaviour. We may therefore expect among-individual variation in behaviour to be greater in absence of supplementary food, when competition for food resources are greater and conditions unfavourable. Individual specialisation may in turn affect reproductive timing and investment and also increase among-individual variation in reproduction if a shift in resource use is associated with using less profitable food items. However, given that higher food availability is often coupled with higher population density, which rises competition levels, among-individual variation may increase in food-supplemented context. The opposite effects of food

availability and density on food resource competition may also cause among-individual variation to be similarly large in both food availability contexts. Alternatively, an increase in niche breath relaxing competition may be achieved by all individuals increasing their range of resource used. In this scenario, among-individual variation would be similar in both low and highly competitive contexts whereas within-individual plasticity would be greater in highly competitive contexts. We may therefore expect within-individual plasticity to differ between food-supplemented and non-supplemented contexts. Again, hypothesising a direction is difficult because of opposite effects of food availability and density on food resource competition.

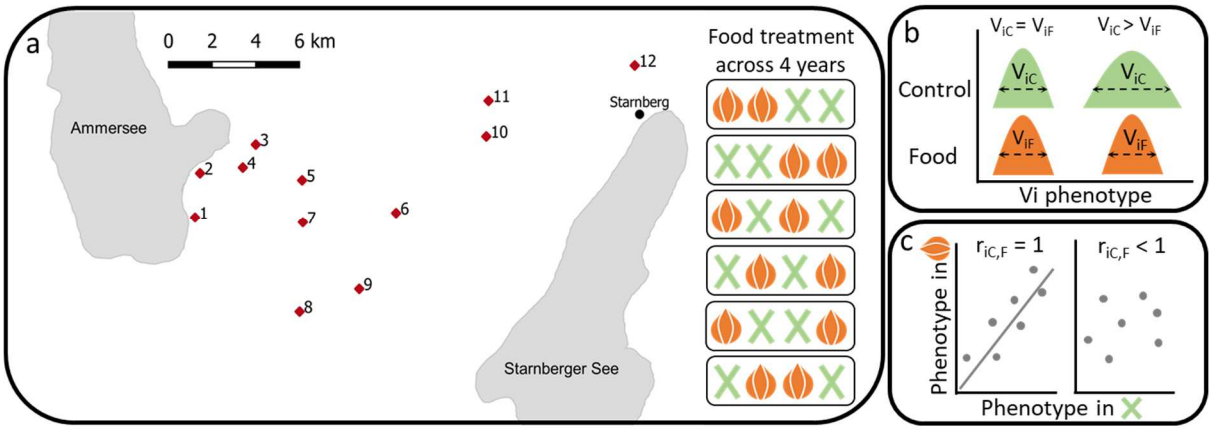


Figure 1. Experimental design of the food treatment. a) Study area with the 12 nest box plots (red diamonds) in which the supplementary feeding experiment was conducted across four consecutive years. Following a partial cross-over design, all possible combinations of supplementary food-control sequences (illustrated with orange sunflower seed icons and green crosses) were achieved randomly across plots, such as all plots were twice treated and twice control. b) This design allowed us to estimate and compare among-individual variances in control and supplementary fed plots. c) It also allowed us to estimate the among-individual cross-context correlations to determine whether the phenotype of the individuals depended on the environmental context (phenotype-by-environment interaction, $r_{iC\text{-}F} < 1$) or not ($r_{iC\text{-}F} = 1$).

MATERIAL AND METHODS

Data collection

Data were collected southern Munich, Germany (47° 97'N, 11° 21'E), in 12 forest plots in 2016 and 2017, and in 11 of these plots in 2018 and 2019 (access to the 12th plot has been withdrawn). Each plot was fitted with 50 nest boxes covering 8-10 ha. Each breeding season (April-July), nest boxes were monitored at least once per week to determine lay date (back-calculated assuming one egg laid per day) and clutch size. Female great tits were caught in their nest box using a spring trap when their nestlings were 10-12 days old, and ringed at first capture. Each bird immediately performed an activity test in a cage that represents a field version of the novel environment test assaying exploration behaviour in the lab (Dingemanse

et al. 2002). We used the field- instead of the laboratory-based assay because it allowed us to score the behaviour of the whole breeding population and avoid sampling biases with respect to behavioural phenotypes (Biro & Dingemanse 2009; Stuber et al. 2013; Mouchet & Dingemanse 2021). The cage (61L × 39W × 40H cm³) was placed before capture at a distance of at least 50m from the focal nest box and consisted in opaque material and a front metal grid. The bird was first placed a few seconds in a small box (11L × 12W × 11H cm³) connected to the cage and darkened with a bag for acclimatization. The observer then opened the sliding door of the holding box and removed the bag to stimulate the bird to fly into the cage without handling. The behaviour was video recorded for the first two minutes the bird entered the cage. Activity scores were calculated summing the number of movements birds made between three perches, three floor sections and six sections on the grid (illustrated in Stuber et al. 2013).

Food supplementation

Supplementary food was provided outside the breeding season continuously from mid-July of a year until end of March of the following year, during four consecutive autumns/winters (2015-2019). In each treated plot, we placed four feeders such that they covered the entire plot. We refilled each of them once per week with ca. 2kg sunflower seeds and placed ca. 300g fat balls in the vicinity. Each year, half of the plots received supplementary food (food-supplemented context), while the other half was unmanipulated (non-supplemented context). Following a partial cross-over design, all plots received each treatment condition (supplementary food or control) twice across the four years following a semi-randomised order (Figure 1). We therefore repeatedly measured behaviour and reproduction within food context in some individuals. Because behaviour and annual reproduction are labile traits, this design forcefully allowed us to disentangle among-individual variation and within-individual plasticity. We estimated selection in food-supplemented and non-supplemented contexts and determined whether selection on behaviour varied with food treatment. We also estimated among-individual variance in behaviour and reproductive traits in both food contexts and tested whether the amount of variance was food-context specific. Finally, we estimated among-individual cross-context correlations to test whether individuals all responded the same way to changes in food availability.

Statistical analyses

To test whether our food manipulation increased breeding densities, we ran a linear mixed-effect model with breeding density calculated as the number of breeding pairs/ha as the

response variable. We fitted food treatment (fitted as 2-level factor), year (4-level factor), and their interaction as fixed effects, and plot as random factor.

We then investigated the effect of food availability on the relationships between the three traits activity score, lay date and clutch size, using a multivariate linear mixed effects model with random intercepts and Gaussian error. We used data only from females because it has previously been shown that, in our population, lay date and clutch size are female traits little influenced by their male partner (Araya-Ajoy et al. 2016). We fitted each focal trait expressed in each food treatment as separate variables (e.g., activity expressed in non-supplemented vs. activity expressed in food-supplemented contexts). Thus, the model had six response variables. This multivariate model was performed in R version 3.5.3 (R Core Team) using the MCMCglmm package (Hadfield 2010). It was run with 3,000,000 iterations, from which we discarded the first 600,000. The chains were stored at intervals of 2000 iterations, resulting in low autocorrelation (<0.1) between stored iterations. The fixed effect prior was normally distributed, with zero mean and large variance of 100. The prior for variances and covariances was non-informative.

As fixed effects, we fitted the covariates age, age squared, and minimum and maximum age because a previous study showed a non-linear effect of age on activity, lay date and clutch size and selective (dis)appearance (Dingemanse et al. 2020). Minimum and maximum age represented the age the individual bred for the first and last time in the dataset. We further fitted the four-level factor year to account for variation in the phenotypic traits due to yearly changes in environmental conditions else than the food treatment, particularly in natural food availability that might interact with our experiment. As random factors, we fitted Individual ID, Plot and the unique combination of Plot and Year (Plot-Year) to estimate, respectively, among-individual, spatial and spatiotemporal variation in the traits. We also estimated the repeatability of each trait.

To test whether the mean values of exploration, lay date and clutch size differed between the treatments, we generated a distribution of the difference between the posterior distributions estimated for a focal trait in food-supplemented vs. non-supplemented context. We then estimated the 89% credible interval (CI) of this distribution as an estimate of uncertainty. The effect was considered biologically important when the interval did not overlap zero. We used 89% CI instead of the usual 95% CI because it is proposed to be more stable for Bayesian posterior distributions (Kruschke 2014; McElreath 2016). We used the same approach to test differences in among- and within-individual variance and repeatability between treatments. Finally, we estimated selection on activity and lay date using clutch size as fitness proxy. We

estimated selection gradients (and 89% CI) using partial regression coefficients obtained from the multivariate model between clutch size and behaviour or lay date and accounting for within-individual variance in the traits (Dingemanse et al. 2021). That is, selection gradients were estimated solely based on among-individual covariances between phenotypic traits and fitness, which avoids underestimation of selection gradients due do within-individual plasticity in labile traits (Dingemanse et al. 2021). Similarly, pairwise correlations cross-traits within-context at the among- and within-individual level and within-trait cross-context only at the among-individual level were considered different from zero when the 89% CI did not overlap zero. We tested whether the cross-context correlation was different from one by using the region of practical equivalence (ROPE), which indicates the percentage of the distribution being in a chosen interval. We set the interval of the region to (0.90-1) and considered a correlation with a ROPE (0.90-1) greater or equal to 5% not being different from one.

We used this multivariate approach to take forward the uncertainty in the estimation of the relationships between traits within contexts and within traits across contexts, and variances and covariances. Because food treatment changed among plots within years and within plots across years, but not within plots within years, an individual was not exposed to both treatments at the same time. Therefore, within-individual cross-context covariances were not estimable, and thus ignored. Similarly, the cross-context covariances at plot-year level were not estimable. By contrast, because some individuals and all plots were exposed to both treatments across the four years of the experiment, among-individual and among-plot cross-context covariances were estimable. From the variances and covariances, we estimated correlations.

RESULTS

Food supplementation from July to March increased the breeding density in the subsequent breeding season from an average density of 2.03 pairs/ha in non-supplemented context to 2.50 pairs/ha in food-supplemented context (Table S1). The dataset used for selection and variance analyses on behaviour and fitness comprised 662 observations from 470 females, thus the average female was recorded 1.40 times. The average lay date (LD) and clutch size (CS) were similar in both food contexts (LD: $mode_F$ = 19.25 (16.76, 20.83), $mode_C$ = 17.36 (15.12, 19.30), $\Delta mode_{F,C}$ (89% CI) = 0.96 (-0.71,3.91); CS: $mode_F$ = 8.59 (7.29, 9.41), $mode_C$ = 7.75 (6.67, 8.82), $\Delta mode_{F,C}$ = 0.82 (-0.62, 1.72), Table S2) . Among- and within-individual variance in lay date and clutch size did not differ across food availability contexts either (Vi LD: σ^2_F = 7.42 (4.19, 12.35), σ^2_C = 9.30 (5.01, 12.90), $\Delta Vi_{F,C}$ = -0.92 (-6.51, 3.88); Vi CS: σ^2_F = 1.06 (0.85, 1.52), σ^2_C = 0.90 (0.71, 1.33), $\Delta Vi_{F,C}$ = 0.24 (-0.29, 0.58); Vw LD: σ^2_F = 14.92 (11.81, 19.32),

σ^2_C = 13.16 (10.63, 17.98), $\Delta Vw_{F,C}$ = 1.17 (-4.95, 2.80); Vw CS: σ^2_F = 0.89 (0.64, 1.18), σ^2_C = 1.06 (0.87, 1.39), $\Delta Vw_{F,C}$ = -0.12 (-0.57, 019); Table S3). The cross-context correlation in lay date differed from 0 but not from 1, implying that a female that laid relatively early (or late) in food-supplemented context also laid relatively early (or late) in non-supplemented context ($ri_{C,F}$ = 0.90 (0.72,0.95), ROPE = 23.75%; Figure 2). However, the cross-context correlation in clutch size differed both from 0 and 1, implying that females responded differently to food availability change: some females produced large or small clutches in both contexts, while other females produced a large clutch in one context and a small clutch in another context ($ri_{C,F}$ = 0.63 (0.35,0.72), ROPE = 0.00%; Figure 2). Combined with similar variance and mean in clutch size across contexts, this finding also implies that a similar number of females switched their relative clutch size in opposite direction. The average activity score was higher in years with supplementary food ($\Delta mode_{F,C}$ = 10.31 (0.98, 14.69), Table S2). Among-individual variance and repeatability in activity were also greater in food-supplemented context (σ^2_F = 209.93 (150.43, 294.86), σ^2_C = 99.51 (61.25, 177.30), ΔVi_{F-C} = 104.60 (22.71, 219.67); Figure 3; Table S3). The cross-context correlation in activity differed from 0 but not from 1, implying that birds all responded similarly to the change in food availability: birds that were relatively more active in food-supplemented context were also relatively more active in non-supplemented context ($ri_{C,F}$ = 0.99 (0.93,1), ROPE = 95.00%; Figure 2). By contrast, the within-context within-individual variation in activity tended to be lower in food-supplemented context (σ^2_F = 181.56 (121.66, 244.24), σ^2_C = 287.37 (229.16, 349.86), ΔVw_{T-C} = -101.49 (-185.72, 2.75); Figure 2; Table S3).

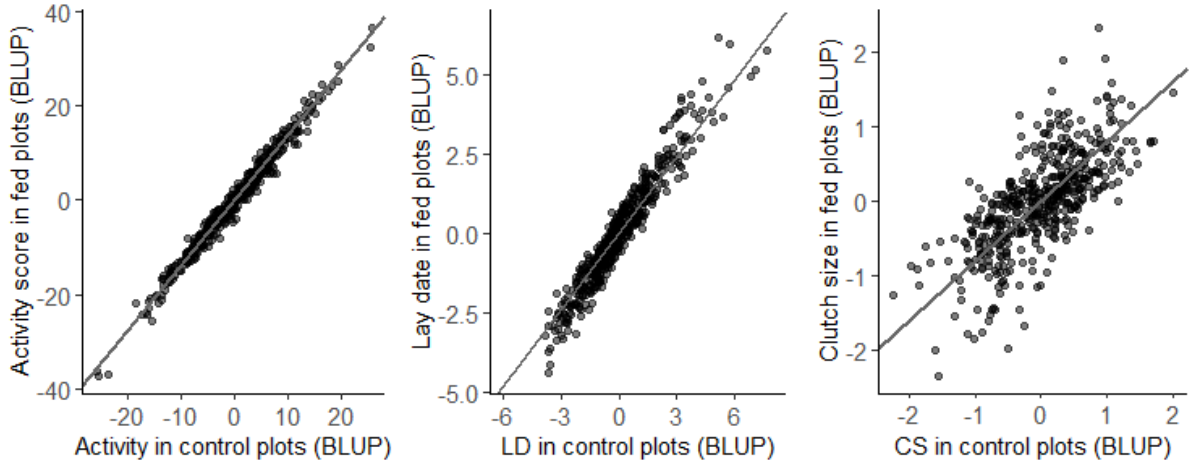


Figure 2. Cross-context correlation in activity score, lay date and clutch size. Cross-context correlations in activity score and lay date were not different from 1, implying individuals kept their phenotypic ranks across food contexts. By contrast, cross-context correlation in clutch size was smaller than 1, implying phenotype-by-environment interactions with respect to food availability.

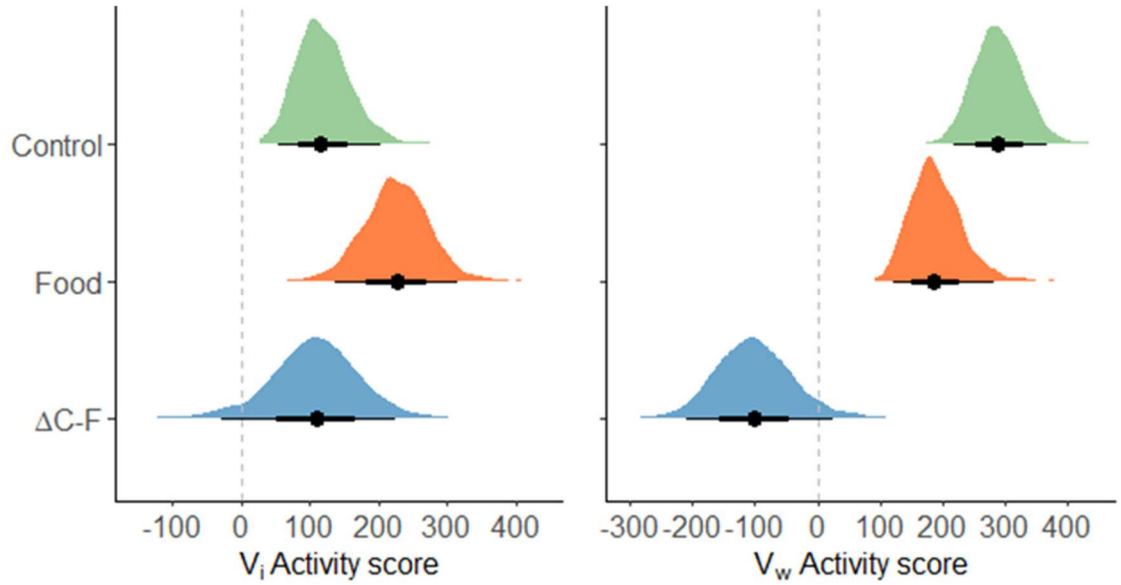


Figure 3. Among and within-individual variance in activity score in control and supplementary fed plots. Among-individual variance (V_i) was greater in years with supplementary food whereas within-individual variance (V_w) tended to be larger in years without supplementary food.

The among-individual cross-trait correlations showed that activity did not predict laying date in food-supplemented context, whereas in non-supplemented context, more active females laid later than less active females (Table 1). The correlations in each food context, however, did not significantly differ from each other ($\Delta Cov_{IF\text{-}C}$ = -0.82 (-15.87, 13.98); Table S4). We may therefore assume that, in both food contexts, more active females laid later, though the relationship was less strong in food-supplemented context. Because the correlation is calculated with the covariance and the variance of each trait, this result aligns with covariances between lay date and activity being similar across food contexts and only the variance in activity increasing in food-supplemented context. Females that laid later tended to lay smaller clutches, and at a similar extent in both food contexts (Table 1, Table S4). However, when females laid at the same date, activity did not predict clutch size in any food context (Table 1). Estimating the selection gradient between activity and clutch size confirmed that activity was not under directional selection in any food context (non-supplemented: β (89% CI) = 0.009 (-0.02, 0.02); food-supplemented: β = -0.01 (-0.08, 0.05); Figure 4). Therefore, selection on activity also did not vary with food availability. Selection tended to favour earlier laying date in both food contexts (non-supplemented: β = -0.04 (-0.13, 0.02); food-supplemented: β = -0.05 (-0.12, 0.03)). The within-individual cross-trait correlations showed that activity correlated with laying date only in non-supplemented context, and in a negative manner (Table 1). The within-individual correlations in each food context significantly differed from each

other ($\Delta Cov_{WF\text{-}C}$ = 21.58 (3.31, 33.15); Table S4). Thus, a female that was more active in a year of low food relatively to another year of low food also laid earlier. This relationship between activity and laying date at the within-individual level was opposite to the among-individual level. That is, in non-supplemented context, more active females laid later compared to less active females, but when a female increased her activity level from one year to the next, she also shifted to lay earlier. A female also plastically decreased her clutch size when laying later in non-supplemented but not in food-supplemented context (Table 1). She did not however change her clutch size with her activity level in any food context (Table 1).

Table 1. Cross-trait within-context correlations between activity, laying date and clutch size at the among- and within-individual level. Pairwise correlations in low (i.e. control) vs. high availability food context (i.e., treated plots) at the among-individual level (respectively, $r_{IC\text{-}C}$ vs. $r_{IT\text{-}T}$) and the within-individual level (respectively, $r_{RC\text{-}C}$ and $r_{RT\text{-}T}$) are printed with their 89% credible interval (89% CI).

	Among-individual		**Within-individual**	
	$r_{IC\text{-}C}$ (89% CI)	$r_{IT\text{-}T}$ (89% CI)	$r_{RC\text{-}C}$ (89% CI)	$r_{RT\text{-}T}$ (89% CI)
Activity-Lay date	0.42 (0.01,0.65)	0.19 (-0.1,0.51)	-0.21 (-0.36,-0.04)	0.08 (-0.10,0.31)
Activity-Clutch size	-0.03 (-0.23,0.24)	-0.19 (-0.31,0.13)	0.01 (-0.12,0.17)	-0.02 (-0.25,0.2)
Lay date-Clutch size	-0.24 (-0.44,0.04)	-0.25 (-0.5,0.05)	-0.13 (-0.35,-0.02)	-0.04 (-0.23,0.14)

DISCUSSION

Heterogeneous selection on behaviour could favour the persistence of among-individual variation in behaviour within populations either by alternately favouring different phenotypes or by favouring different phenotypes in different locations (Frank & Slatkin 1990; Wolf & Weissing 2010). Heterogeneous selection on behaviour could be driven by spatiotemporal variation in food availability as this ecological factor strongly affects survival and reproductive success, and individuals often differ in their foraging tactics and competitive abilities. In contrast with previous findings, fecundity selection on activity behaviour did not vary with our experimental manipulation of food availability outside the breeding season (Dingemanse et al. 2004; Boon et al. 2007; Le Cœur et al. 2015).

Our supplementary food experiment mimicked beech masting, an important food source for great tits outside the breeding season that positively affects their survival (Perdeck et al. 2000). When food is scarce and clumped, that is when competition for food resources is high, more competitive, active individuals may have greater access to food and survive the bad season better. This assumption could explain previous findings of a positive link between exploration behaviour and survival of female great tits in non-beech mast years (Dingemanse et al. 2004). However, the female population-average activity score and the among-individual

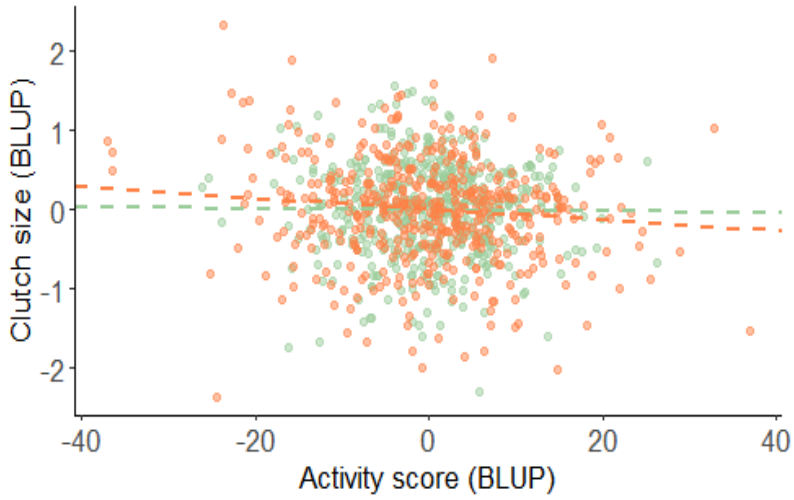


Figure 4. Selection gradients on activity score in low and high food availability context do not differ. Dots represent the best linear unbiased predictors (BLUPs) for clutch size and activity score and dashed lines the selection gradients in each food treatment (control: green, supplementary food: orange)

variance were greater in years with (instead of without) supplementary food, suggesting higher survival, recruitment, and/or immigration rate of more active phenotypes in years of abundant food. This result aligns with differences in resource distribution between low and high food environments that may induce differences in resource acquisition between behavioural phenotypes. Indeed, more active individuals rely more on social cues and require more energy (Groothuis & Carere 2005); this may be disadvantageous in low food environment where population density, thus the number of individuals to scrounge from, is low and food patches may be distant. Indeed, in non-supplemented context, more active females laid later, and consequently, tended to lay smaller clutches, than less active females. In food-supplemented context, more active females may be advantaged over less active females because their competitive advantage and use of social cues may be beneficial when food abounds, and density increases throughout the winter. More active females may, in turn, be disadvantaged at the onset/during the subsequent breeding season when breeding density is high if costs of competitiveness are great due to risk of injuries and energy/time invested in agonistic interactions (Verbeek et al. 1999; Wolf et al. 2007; Careau & Garland 2012; Niemelä & Dingemanse 2018; Mathot et al. 2019). This may explain why more active females tended to also lay later than less active females in food-supplemented. More active females did not compensate late laying by laying larger clutches when laying at the same date than less active females. Therefore, though direct food-dependent selection on activity behaviour was neutral, more active females overall tended to lay smaller clutches in both food contexts. Altogether, these findings imply that food-dependent selection on activity may be counteracted by density-

dependence, which could explain a lack of heterogeneous food-dependent selection on activity behaviour. This may also explain that heterogeneous selection on exploration behaviour in West European great tit populations was not driven by density-dependent selection (Mouchet et al. 2021a). Indirect selection via lay date also appeared to counteract direct selection and favour the same phenotypes in both food contexts. Thus, for variation to be maintained, more active females may be favoured by other conditions than food availability, for example increased predation risk (Abbey-Lee & Dingemanse 2019). Opposite selection patterns induced by multiple (covarying) ecological factors may cause behavioural phenotypes to reap the same reproductive output within breeding seasons and be maintained within populations. To better understand how behavioural phenotypes coexist, future studies should estimate selection on behaviour driven by multiple interacting ecological factors, such as food availability, density and predation risk, simultaneously.

Interestingly, females that increased their activity across non-supplemented contexts laid earlier than when less active, which is opposite to the among-individual pattern. By contrast, in food-supplemented context, females did not change their laying date with a change in their activity score. Given that within-individual correlations where estimated based on data from different years, other ecological factors than food availability may have changed across years. Thus, activity and lay date may be either causally linked or both affected by a third factor, with food availability outside the breeding season being a key mediating factor. Temperature in the couple of months before the start of the breeding season has been shown to be an important factor affecting lay date of great tits (Visser et al. 2009). Therefore, across years of low food availability, females may be able to increase their activity when, for example, winter temperatures are milder, which may also allow them to lay earlier. By contrast, continuous access to food in food-supplemented context may allow females to lay at the same date irrespective of other ecological factors and their activity level. This explanation aligns with the finding that females also plastically decreased their clutch size with later laying date only in non-supplemented context. Abundant winter food resources thus compensate the negative effect of laying late.

The absence of heterogeneous food-dependent selection on activity behaviour may be due to variable feeding conditions affecting selection not on mean behaviour but on plasticity. Ecological conditions intensifying competition for resources may cause individuals to either express wider behavioural plasticity or modify their behaviour to use different resources relatively to others in the population (social niche specialisation) (Bolnick et al. 2007; Bergmüller & Taborsky 2010). The greater among-individual variance in activity behaviour in

food-supplemented context aligns with the social niche hypothesis where individuals specialise in a social role to release competition with conspecifics (Bergmüller & Taborsky 2010). This finding suggests that competition induced by increased population density in food-supplemented context is stronger than competition induced by low food availability. Altogether with increased mean activity and stable individual ranks, the increased among-individual variation also implies that relatively more active individuals were more plastic. In non-supplemented context, food patches may be unevenly distributed, rending the environment more heterogeneous. These conditions may not allow individuals to acquire the energy required to express higher levels of activity or may decrease the benefits of foraging tactics associated with high activity levels, such as superficially explore the environment and relying more on social cues (Verbeek et al. 1994; Groothuis & Carere 2005). Alternatively, food availability may drive selection through other fitness metrics than clutch size. More active females may be favoured in food-supplemented context through higher survival and/or offspring recruitment rate instead of clutch size. As mentioned above, this is what suggests the higher population-mean activity score in food-supplemented context. The simultaneous increase in among-individual variance in behaviour suggests viability selection may be weaker in years of abundant food and stronger in years of low food.

The lower within-individual variance in food-supplemented context also aligns with individual specialisation in social niches and density, rather than food availability, imposing greater competition. When competition is relatively relaxed, individuals may use overlapping niches (Bolnick et al. 2003, 2007). Individuals thus may use a wider range of resources, thereby expanding their own niche. This expansion may require adjusting activity levels. By contrast, when competition is intense, individuals may use distinct niches, which they may achieve by narrowing the array of resources used. This niche shrinkage may result in individuals expressing narrow ranges of activity levels adjusted to the resources used. The difference in within-individual variation across food contexts may also result from changes in environmental predictability. If the environment is highly variable and heterogeneous, such as when food is scarce, a female might need to adapt her behaviour more than when the environment is more predictable, such as when winter food abounds. If this hypothesis holds, given that the feeding conditions pre-breeding and during breeding might differ, and that we assayed activity during the breeding season, the feeding conditions pre-breeding would determine the plasticity level a female shows during the breeding season.

Though fecundity selection on activity behaviour did not vary with food availability, females responded differently to a change in food availability, irrespective of their activity.

Some females laid a similar clutch size in both food contexts, while others laid relatively large clutches in one food context and relatively small clutches in the other food context. This finding suggests that other phenotypic traits than activity behaviour may be under food-dependent selection. It may be other behaviours, such as aggressiveness or boldness, or physiological or morphological traits (Grant & Grant 2002; Le Cœur et al. 2015). A change in clutch size was not due to a change in lay date as females laid at a relatively similar date in both food contexts. Thus, winter food availability appears to determine clutch size more strongly than lay date, though only for some females. Interestingly, the average lay date and clutch size were similar in both food contexts. Previous studies showed density-dependent reproduction, with birds laying earlier and smaller clutch sizes at higher densities (Both 1998; Wilkin et al. 2006; Ahola et al. 2009). Though our food manipulation increased density, birds did not adjust their reproductive timing and investment.

Variation in food availability is often suggested to drive heterogeneous selection on behaviour and thus be a factor favouring the coexistence of personalities. Our study manipulating winter food availability rejects this hypothesis for activity behaviour in novel environments. However, counteracting density-dependent selection pressures may be at play and cancel out the effect of food-dependent selection. Disentangling the effects of covarying food availability and population density on personality-related selection is an exciting future challenge. Females reproductively responded differently to variation in food availability, suggesting other behavioural phenotypes may be under heterogeneous food-dependent selection. Personality-related food-dependent selection may also act through survival or recruitment rate, by favouring more active females in years of abundant food. Our work suggests multivariate actions and sources of selection favouring the adaptive maintenance of individual behavioural variation.

ACKNOWLEDGMENTS

We thank Maria Moiron, Alexander Hutfluss, field assistants and students of the Behavioral Ecology group (LMU) for helping to collect the data. All procedures complied with the guidelines from the District Government of Upper Bavaria (Regierung von Oberbayern) for Animal care ROB-55.2-2532.Vet_02-17-215. This work was supported by the German Science Foundation (grant no. DI 1694/1-1) and by the Max-Planck Society (grant to N.J.D. 2012-2017). A.M. was part of the International Max-Planck Research School for Organismal Biology.

REFERENCES

Abbey-Lee, R.N. & Dingemanse, N.J. (2019). Adaptive individual variation in phenological responses to perceived predation levels. Nat. Commun., 10, 1–8.

Ahola, M.P., Laaksonen, T., Eeva, T. & Lehikoinen, E. (2009). Great tits lay increasingly smaller clutches than selected for: A study of climate- and density-related changes in reproductive traits. J. Anim. Ecol., 78, 1298–1306.

Amy, M., Sprau, P., De Goede, P. & Naguib, M. (2010). Effects of personality on territory defence in communication networks: A playback experiment with radio-tagged great tits. Proc. R. Soc. B Biol. Sci., 277, 3685–3692.

Araya-Ajoy, Y.G., Kuhn, S., Mathot, K.J., Mouchet, A., Mutzel, A., Nicolaus, M., et al. (2016). Sources of (co)variation in alternative siring routes available to male great tits (Parus major). Evolution (N. Y)., 70, 2308–2321.

Balen, J.H. Van. (1980). Population Fluctuations of the Great Tit and Feeding Conditions in Winter. Ardea, 68, 143–164.

Barber, I. & Dingemanse, N.J. (2010). Parasitism and the evolutionary ecology of animal personality. Philos. Trans. R. Soc. B Biol. Sci., 365, 4077–4088.

Bergmüller, R. & Taborsky, M. (2010). Animal personality due to social niche specialisation. Trends Ecol. Evol., 25, 504–511.

Biro, P.A. & Dingemanse, N.J. (2009). Sampling bias resulting from animal personality. Trends Ecol. Evol., 24, 63–66.

Biro, P.A. & Stamps, J.A. (2009). Are animal personality traits linked to life-history productivity? Trends Ecol. Evol., 24, 361–368.

Bolnick, D.I., Svanbäck, R., Araújo, M.S. & Persson, L. (2007). Comparative support for the niche variation hypothesis that more generalized populations also are more heterogeneous. Proc. Natl. Acad. Sci. U. S. A., 104, 10075–10079.

Bolnick, D.I., Svanbäck, R., Fordyce, J.A., Yang, L.H., Davis, J.M., Hulsey, C.D., et al. (2003). The ecology of individuals: Incidence and implications of individual specialization. Am. Nat., 161, 1–28.

Boon, A.K., Réale, D. & Boutin, S. (2007). The interaction between personality, offspring fitness and food abundance in North American red squirrels. Ecol. Lett., 10, 1094–1104.

Both, C. (1998). Experimental evidence for density dependence of reproduction in great tits. J. Anim. Ecol., 67, 667–674.

Byers, D.L. (2005). Evolution in heterogeneous environments and the potential of maintenance of genetic variation in traits of adaptive significance. Genetica, 123, 107–124.

Careau, V. & Garland, T. (2012). Performance, personality, and energetics: Correlation, causation, and mechanism. Physiol. Biochem. Zool., 85, 543–571.

Le Cœur, C., Thibault, M., Pisanu, B., Thibault, S., Chapuis, J.L. & Baudry, E. (2015). Temporally fluctuating selection on a personality trait in a wild rodent population. Behav. Ecol., 26, 1285–1291.

Cote, J., Dreiss, A. & Clobert, J. (2008). Social personality trait and fitness. Proc. R. Soc. B Biol. Sci., 275, 2851–2858.

Dall, S.R.X., Houston, A.I. & McNamara, J.M. (2004). The behavioural ecology of personality: Consistent individual differences from an adaptive perspective. Ecol. Lett., 7, 734–739.

Dammhahn, M., Dingemanse, N.J., Niemelä, P.T. & Réale, D. (2018). Pace-of-life syndromes: a framework for the adaptive integration of behaviour, physiology and life history. Behav. Ecol. Sociobiol., 72, 1–8.

Dingemanse, N.J., Araya-Ajoy, Y.G. & Westneat, D.F. (2021). Most published selection gradients are underestimated: Why this is and how to fix it. Evolution (N. Y)., 806–818.

Dingemanse, N.J., Both, C., Drent, P.J., Van Oers, K. & Van Noordwijk, A.J. (2002). Repeatability and heritability of exploratory behaviour in great tits from the wild. Anim. Behav., 64, 929–938.

Dingemanse, N.J., Both, C., Drent, P.J. & Tinbergen, J.M. (2004). Fitness consequences of avian personalities in a fluctuating environment. Proc. R. Soc. B Biol. Sci., 271, 847–852.

Dingemanse, N.J., Bouwman, K.M., van de Pol, M., van Overveld, T., Patrick, S.C., Matthysen, E., et al. (2012). Variation in personality and behavioural plasticity across four populations of the great tit Parus major. J. Anim. Ecol., 81, 116–126.

Dingemanse, N.J. & De Goede, P. (2004). The relation between dominance and exploratory behavior is context-dependent in wild great tits. Behav. Ecol., 15, 1023–1030.

Dingemanse, N.J., Kazem, A.J.N., Réale, D. & Wright, J. (2010). Behavioural reaction norms: animal personality meets individual plasticity. Trends Ecol. Evol., 25, 81–89.

Dingemanse, N.J., Moiron, M., Araya-Ajoy, Y.G., Mouchet, A. & Abbey-Lee, R.N. (2020). Individual variation in age-dependent reproduction: Fast explorers live fast but senesce young? J. Anim. Ecol., 89, 601–613.

Dingemanse, N.J. & Wolf, M. (2010). Recent models for adaptive personality differences: A review. Philos. Trans. R. Soc. B Biol. Sci., 365, 3947–3958.

Felsenstein, J. (1976). The theoritical population genetics of variable selection and migration. Annu. Rev. Genet., 10, 253–280.

Foster, S.A. & Endler, J.A. (1999). Geographic variation in behavior. Perspectives on evolutionary mechanisms. Oxford University Press, New York.

Frank, S.A. & Slatkin, M. (1990). Evolution in a Variable Environment. Am. Nat., 136, 244–260.

Le Galliard, J.F., Paquet, M. & Mugabo, M. (2015). An experimental test of density-dependent selection on temperament traits of activity, boldness and sociability. J. Evol. Biol., 28, 1144–1155.

Gillespie, J.H. & Turelli, M. (1989). Genotype-environment interactions and the maintenance of polygenic variation. Genetics, 121, 129–138.

Grant, P.R. & Grant, B.R. (2002). Unpredictable Evolution in a 30Year Study Darwin Finches, 296, 707–712.

Groothuis, T.G.G. & Carere, C. (2005). Avian personalities: Characterization and epigenesis. Neurosci. Biobehav. Rev., 29, 137–150.

Grøtan, V., Saether, B.-E., Engen, S., van Balen, J.H., Perdeck, A.C. & Visser, M.E. (2009). Spatial and temporal variation in the relative contribution of density dependence, climate variation and migration to fluctuations in the size of great tit populations. J. Anim. Ecol., 78, 447–459.

Hadfield, J.D. (2010). MCMCglmm: MCMC Methods for Multi-Response GLMMs in R. J. Stat. Softw., 33, 1–22.

Haldane, J.B.S. & Jayakar, S.D. (1963). Polymorphism due to selection of varying direction. J. Genet., 58, 237–242.

Hedrick, P.W. (1976). Genetic variation in a heterogeneous environment. II. Temporal heterogeneity and directional selection. Genetics, 84, 145–157.

Hoffmann, A.A. & Merilä, J. (1999). Heritable variation and evolution under favourable and unfavourable conditions. Trends Ecol. Evol., 14, 96–101.

Hollander, F.A., Van Overveld, T., Tokka, I. & Matthysen, E. (2008). Personality and nest defence in the great tit (Parus major). Ethology, 114, 405–412.

Kruschke, J.K. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan, second edition. Doing Bayesian Data Anal. A Tutor. with R, JAGS, Stan, Second Ed.

Lack, D. (1954). The natural regulation of animal numbers. Oxford Univ. Press.

Lack, D. (1964). A Long-Term Study of the Great Tit (Parus major). J. Anim. Ecol., 159–173.

Lande, R. (1976). Natural Selection and Random Genetic Drift in Phenotypic Evolution. Evolution (N. Y)., 30, 314.

Mathot, K.J., Dingemanse, N.J. & Nakagawa, S. (2019). The covariance between metabolic rate and behaviour varies across behaviours and thermal types: meta-analytic insights. Biol. Rev., 94, 1056–1074.

McElreath, R. (2016). Statistical Rethinking: A Bayesian Course with Examples in R and Stan. 1st Editio. Boca Raton.

Moiron, M., Araya-Ajoy, Y.G., Mathot, K.J., Mouchet, A. & Dingemanse, N.J. (2019). Functional relations between body mass and risk-taking behavior in wild great tits. Behav. Ecol., 30, 617–623.

Moiron, M., Laskowski, K.L. & Niemelä, P.T. (2020). Individual differences in behaviour explain variation in survival: a meta-analysis. Ecol. Lett., 23, 399–408.

Mouchet, A., Cole, E.F., Matthysen, E., Nicolaus, M., Quinn, J.L. & Roth, A.M. (2021). Heterogeneous selection on exploration behavior within and among West European populations of a passerine bird. Proc. Natl. Acad. Sci., 118, 1–6.

Mouchet, A. & Dingemanse, N.J. (2021). A quantitative genetics approach to validate lab-versus field-based behavior in novel environments. Behav. Ecol., 1–9.

Mutzel, A., Dingemanse, N.J., Araya-Ajoy, Y.G. & Kempenaers, B. (2013). Parental provisioning behaviour plays a key role in linking personality with reproductive success. Proc. R. Soc. B Biol. Sci., 280, 1–9.

Nicolaus, M., Tinbergen, J.M., Ubels, R., Both, C. & Dingemanse, N.J. (2016). Density fluctuations represent a key process maintaining personality variation in a wild passerine bird. Ecol. Lett.

Niemelä, P.T. & Dingemanse, N.J. (2018). Meta-analysis reveals weak associations between intrinsic state and personality. Proc. R. Soc. B Biol. Sci., 285, 1–7.

Overveld, T. Van & Matthysen, E. (2010). Personality predicts spatial responses to food manipulations in free-ranging great tits (Parus major). Biol. Lett., 6, 187–190.

Perdeck, A.C., Visser, M.E. & van Balen, J.H. (2000). Great tit Parus major survival and the beech-crop cycle. Ardea, 88 (1), 99–108.

Perrins, C.M. (1965). Population fluctuations and clutch size in the great tit, Parus major L. J. Anim. Ecol., 34, 601–647.

Quinn, J.L., Patrick, S.C., Bouwhuis, S., Wilkin, T.A. & Sheldon, B.C. (2009). Heterogeneous selection on a heritable temperament trait in a variable environment. J. Anim. Ecol., 78, 1203–1215.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rollins, R.E., Mouchet, A., Margos, G., Chitimia-Dobler, L., Fingerle, V., Becker, N.S., et al. (2021). Repeatable differences in exploratory behaviour predict tick infestation probability in wild great tits. Behav. Ecol. Sociobiol., 75.

Royauté, R., Anderson Berdal, M., Garrison, C.R. & Dochtermann, N.A. (2018). Paceless life? A meta-analysis of the pace-of-life syndrome hypothesis. Behav. Ecol. Sociobiol., 72, 1–10.

Siepielski, A.M., Dibattista, J.D. & Carlson, S.M. (2009). It's about time: The temporal dynamics of phenotypic selection in the wild. Ecol. Lett., 12, 1261–1276.

Siepielski, A.M., Gotanda, K.M., Morrissey, M.B., Diamond, S.E., Dibattista, J.D. & Carlson, S.M. (2013). The spatial patterns of directional phenotypic selection. Ecol. Lett., 16, 1382–1392.

Smith, B.R. & Blumstein, D.T. (2008). Fitness consequences of personality: A meta-analysis. Behav. Ecol., 19, 448–455.

Stuber, E.F., Araya-Ajoy, Y.G., Mathot, K.J., Mutzel, A., Nicolaus, M., Wijmenga, J.J., et al. (2013). Slow explorers take less risk: A problem of sampling bias in ecological studies. Behav. Ecol., 24, 1092–1098.

van Valen, L. (1965). Morphological Variation and Width of Ecological Niche. Am. Nat., 99, 377–390.

Verbeek, M.E.M., Boon, A. & Drent, P.J. (1996). Exploration, aggressive behaviour and dominance in pair-wise confrontations of juvenile male great tits. Behaviour, 133, 945–963.

Verbeek, M.E.M., Drent, P.J. & Wiepkema, P.R. (1994). Consistent individual differences in early exploratory behaviour of male great tits. Anim. Behav., 48, 1113–1121.

Verbeek, M.E.M., De Goede, P., Drent, P.J. & Wiepkema, P.R. (1999). Individual behavioural characteristics and dominance in aviary groups of great tits. Behaviour, 136, 23–48.

Visser, M.E., Holleman, L.J.M. & Caro, S.P. (2009). Temperature has a causal effect on avian timing of reproduction. Proc. R. Soc. B Biol. Sci., 276, 2323–2331.

Wilkin, T.A., Garant, D., Gosler, A.G. & Sheldon, B.C. (2006). Density effects on life-history traits in a wild population of the great tit Parus major: Analyses of long-term data with GIS techniques. J. Anim. Ecol., 75, 604–615.

Wolf, M., Van Doorn, G.S., Leimar, O. & Weissing, F.J. (2007). Life-history trade-offs favour the evolution of animal personalities. Nature, 447, 581–584.

Wolf, M. & Weissing, F.J. (2010). An explanatory framework for adaptive personality differences. Philos. Trans. R. Soc. B Biol. Sci., 365, 3959–3968.

Wolf, M. & Weissing, F.J. (2012). Animal personalities: Consequences for ecology and evolution. Trends Ecol. Evol., 27, 452–461.

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

Table S1. Effect of supplementary food on breeding density. Estimates with 95% credible interval (95%CI) of fixed (β) and random (σ^2) effects of a linear mixed-effect model fitted with random intercepts and Gaussian error. Breeding density was calculated as the number of breeding pairs per hectare (pairs/ha) within each nest box plot. Breeding year was added as fixed effect to account for temporal variation in breeding density due to other factors than the food manipulation. Plot was fitted as random factor to account for spatial variation (among-plot variation) in breeding density. Supplementary food was provided outside the breeding season (from July of a year till end of March of the following year, see Main text for details) for 4 consecutive years.

Fixed Effects	β (95%CI)
Intercept	2.03 (1.55, 2.50)
Food Treatment	0.47 (0.02, 0.94)
BroodYear2017	-0.01 (-0.43, 0.41)
BroodYear2018	-0.12 (-0.58, 0.36)
BroodYear2019	0.91 (0.48, 1.31)
Food Treatment × BroodYear2017	-0.40 (-1.08, 0.22)
Food Treatment × BroodYear2018	-0.23 (-0.96, 0.47)
Food Treatment × BroodYear2019	-0.64 (-1.38, 0.02)
Random Effects	σ^2 (95%CI)
Plot	0.38 (0.26, 0.50)

Table S2. Average activity score, laying date and clutch size in low (control) and high (treated) food availability contexts. Results of a multivariate linear mixed effects model fitted with random intercepts and Gaussian error. Each trait in each food context was fitted as response variable; thus, the model had six response variables. Random factors were Individual ID, Plot-Year and Plot. We present estimates of fixed effects with 95% credible interval (β (95%CI)) for activity score, lay date and clutch size in low (control) and high (treated) food availability context. Other fixed effects were fitted to account for changes in behaviour and reproductive investment with age and years. Variance estimates are in Table S3.

	Control	**Treated**	**Difference in mean**
Fixed Effects	mode (95%CI)	mode (95%CI)	$\Delta mode_{T-C}$ (89%CI)
Intercept			
Activity	58.55 (52.71,65.77)	67.09 (61.29,72.14)	10.31 (0.98,14.69)
Lay date	17.36 (15.12,19.30)	19.25 (16.76,20.83)	0.96 (-0.71,3.91)
Clutch size	7.75 (6.67,8.82)	8.59 (7.29,9.41)	0.82 (-0.62,1.72)
Age			
Activity	-0.42 (-7.73,5.13)	-5.76 (-10.65,0.12)	-2.55 (-10.84,2.87)
Lay date	-1.63 (-3.22,-0.13)	-2.91 (-4.36,-1.53)	-1.23 (-2.95,0.57)
Clutch size	0.42 (0.1,1.05)	0.5 (0.09,0.87)	-0.17 (-0.59,0.41)
Age^2			
Activity	-1.5 (-2.96,0.08)	0.49 (-0.77,1.46)	2.15 (0.20,3.31)
Lay date	0.46 (0.10,0.83)	0.83 (0.53,1.11)	0.41 (0.01,0.77)
Clutch size	-0.14 (-0.27,-0.04)	-0.12 (-0.18,-0.01)	0.08 (-0.04,0.18)
Min age			
Activity	9.94 (2.96,15.08)	6.06 (0.00,12.82)	-2.70 (-9.50,3.53)
Lay date	3.25 (1.66,4.81)	1.53 (0.02,3.16)	-2.28 (-3.24,0.04)
Clutch size	-0.41 (-0.81,0.09)	-0.6 (-1.08,-0.08)	-0.27 (-0.69,0.37)
Max age			
Activity	2.77 (-0.81,5.46)	2.25 (-1.09,5.11)	-1.23 (-3.86,2.96)
Lay date	-0.65 (-1.60,0.07)	-0.51 (-1.31,0.22)	0.06 (-0.67,1.10)
Clutch size	-0.02 (-0.24,0.24)	-0.08 (-0.31,0.14)	-0.17 (-0.37,0.16)
Year 2017			
Activity	13.83 (5.04,18.81)	8.8 (2.68,13.80)	-3.53 (-11.77,3.35)
Lay date	-7.99 (-10.15,-5.49)	-8.13 (-10.6,-6.01)	-0.58 (-3.49,1.94)
Clutch size	0.55 (-0.78,1.62)	-0.52 (-1.59,0.69)	-1.19 (-2.29,0.43)
Year 2018			
Activity	10.54 (2.34,17.42)	0.53 (-5.21,7.06)	-9.06 (-16.65,-1.37)
Lay date	4.08 (1.96,6.82)	3.25 (1.09,5.85)	-1.05 (-3.46,2.18)
Clutch size	0.83 (-0.69,1.99)	0.59 (-0.9,1.54)	-0.20 (-1.63,1.26)
Year 2019			
Activity	9.78 (2.8,17.29)	1.41 (-5.36,8.77)	-9.59 (-16.83,-0.39)
Lay date	-2.16 (-5.08,-0.05)	-5.11 (-7.52,-2.41)	-2.88 (-5.60,0.16)
Clutch size	-0.49 (-1.63,0.90)	-1.02 (-2.38,0.30)	-1.02 (-2.07,0.88)

Table S3. Among- and within-individual variance, and repeatability in activity behaviour, lay date and clutch size in low (control) and high (treated) food availability context. Results of a multivariate linear mixed effects model fitted with random intercepts and Gaussian error. Each trait in each food context was fitted as response variable; thus, the model had six response variables. We present variance estimates (σ^2) of each trait in each food context and the difference in variance between food contexts (treated-control, $\Delta\sigma^2_{T\text{-}C}$) with 89% credible interval (89%CI). Other random factors were fitter to account for spatial (Plot) and spatiotemporal (Plot-Year) variation in behaviour and reproduction. We also present estimates of among-individual repeatability in activity, lay date and clutch size in each food context and difference in repeatability between food contexts. Fixed effects estimates are presented in Table S2.

	Control	**Treated**	**Difference in variance**
Random effects	σ^2 (89%CI)	σ^2 (89%CI)	$\Delta\sigma^2_{T\text{-}C}$ (89%CI)
Individual (V_I)			
Activity	99.51 (61.25,177.30)	209.93 (150.43,294.86)	104.60 (22.71,219.67)
Lay Date	9.30 (5.01,12.90)	7.42 (4.19,12.35)	-0.92 (-6.51,3.88)
Clutch Size	0.90 (0.71,1.33)	1.06 (0.85,1.52)	0.24 (-0.29,0.58)
Plot-Year (V_{PY})			
Activity	1.46 (0.53,11.83)	1.41 (0.45,5.75)	-0.35 (-11.92,4.29)
Lay Date	0.81 (0.45,2.75)	1.16 (0.46,3.21)	0.49 (-1.69,2.40)
Clutch Size	0.60 (0.34,1.15)	0.62 (0.35,1.17)	-0.07 (-0.64,0.67)
Plot (V_{Plot})			
Activity	1.73 (0.59,7.22)	1.59 (0.57,5.60)	-0.04 (-6.37,4.15)
Lay Date	2.32 (0.90,6.14)	1.8 (0.59,5.29)	-0.53 (-4.95,2.80)
Clutch Size	0.85 (0.41,1.81)	0.86 (0.41,1.8)	-0.11 (-1.18,1.03)
Residual (V_R)			
Activity	287.37 (229.16,349.86)	181.56 (121.66,244.24)	-101.49 (-185.72,2.75)
Lay Date	13.16 (10.63,17.98)	14.92 (11.81,19.32)	1.17 (-4.32,5.86)
Clutch Size	1.06 (0.87,1.39)	0.89 (0.64,1.18)	-0.12 (-0.57,0.19)
Total V_P			
Activity	408.42 (368.87,468.29)	409.73 (366.00,473.89)	-5.97 (-63.49,72.87)
Lay Date	28.71 (24.57,32.95)	28.3 (24.50,32.52)	-1.86 (-5.84,5.52)
Clutch Size	3.70 (2.91,4.86)	3.62 (2.97,4.90)	-0.33 (-1.56,1.34)
Repeatability	R ($V_I/(V_I+V_R)$) (89%CI)	R ($V_I/(V_I+V_R)$) (89%CI)	Difference in R ($\Delta R_{T\text{-}C}$; 89%CI)
Individual			
Activity	0.50 (0.15,0.42)	0.51 (0.40,0.69)	0.25 (0.06,0.48)
Lay Date	0.43 (0.24,0.54)	0.28 (0.18,0.49)	-0.02 (-0.23,0.17)
Clutch Size	0.47 (0.35,0.59)	0.59 (0.43,0.68)	0.10 (-0.09,0.24)

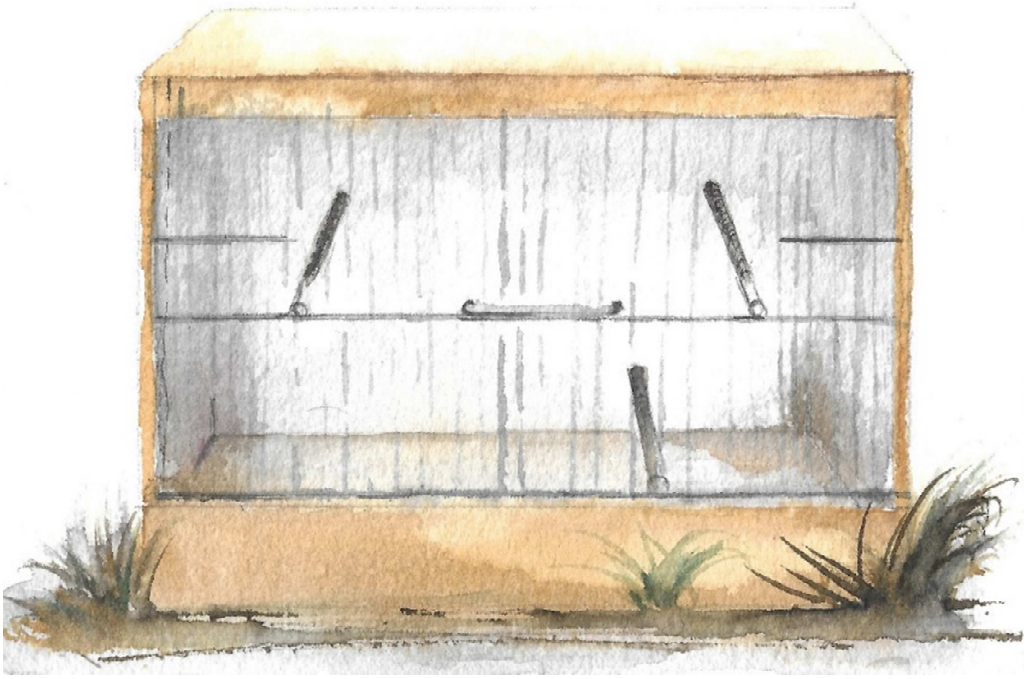
Table S4. Among- and within-individual cross-trait within-context covariance matrices. Results of a multivariate linear mixed effects model fitted with random intercepts and Gaussian error. Each trait in each food context was fitted as response variable; thus, the model had six response variables. Estimates for fixed and random effects are in Table S2 and S3 respectively. We present estimates for the covariance between activity behaviour and lay date (LD), activity and clutch size (CS), and lay date and clutch size in each food context (control: $Cov_{C\text{-}C}$ versus treated: $Cov_{T\text{-}T}$) and the difference in covariances between food contexts ($\Delta Cov_{T\text{-}C}$). All estimates are presented with their 89% credible interval (89%CI).

Among-individual level	$cov_{IC\text{-}C}$ (89%CI)	$cov_{IT\text{-}T}$ (89%CI)	$\Delta Cov_{IT\text{-}C}$
Activity-Lay date	7.42 (0.2,20.44)	6.68 (-2.79,22.47)	-0.82 (-15.87, 13.98)
Activity-Clutch size	-0.59 (-2.38,2.51)	-2.61 (-5.3,1.73)	-1.04 (-5.40, 2.22)
Lay date-Clutch size	-0.64 (-1.31,0.18)	-0.64 (-1.4,0.23)	-0.006 (-1.01, 0.99)
Within-individual level	$cov_{IC\text{-}C}$ (89%CI)	$cov_{IT\text{-}T}$ (89%CI)	$\Delta Cov_{IT\text{-}C}$
Activity-Lay date	-11.77 (-23.21,-2.41)	3.98 (-5.22,17.29)	21.58 (3.31, 33.15)
Activity-Clutch size	-0.46 (-2.3,2.92)	-0.23 (-3.09,2.89)	1.15 (-4.70, 3.10)
Lay date-Clutch size	-0.86 (-1.46,-0.04)	-0.34 (-0.91,0.53)	0.49 (-0.40, 1.58)

Chapter 4

A quantitative genetics approach to validate lab- versus field-based behavior in novel environments

Published in Behavioral Ecology (2021), **32**
https://doi.org/10.1093/beheco/arab059



A quantitative genetics approach to validate lab- versus field-based behavior in novel environments

Alexia Mouchet, Niels J. Dingemanse

Abstract

Conclusions about the adaptive nature of repeatable variation in behavior (i.e., "personality") are often derived from laboratory-based assays. However, the expression of genetic variation differs between laboratory and field. Laboratory-based behavior might not predict field-based behavior thus, cross-context validation is required. We estimated the cross-context correlation between behavior expressed by wild great tits (Parus major) in established laboratory versus field novel environment assays. Both assays have been used as proxies for 'exploration tendency'. Behavior in both contexts had similar repeatability ($R = 0.35$ vs. 0.37) but differed in heritability ($h^2 = 0.06$ vs. 0.23), implying differences in selection pressures. Unexpectedly, there was no cross-context correlation. Laboratory- and field-based behavior thus reflected expressions of two distinct underlying characters. Post hoc simulations revealed that sampling bias did not explain the lack of correlation. Laboratory-based behavior may reflect fear and exploration, but field-based behavior may reflect escape behavior instead, though other functional interpretations cannot be excluded. Thus, in great tits, activity expressed in laboratory vs. field novel environment assays are modulated by multiple quasi-independent characters. The lack of cross-context correlation shown here may also apply to other setups, other repeatable behaviors and other taxa. Our study thus implies care should be taken in labeling behaviors prior to firm validation studies.

Keywords: animal personality, cross-context consistency, heritability, repeatability, exploration behavior, risk-taking behavior

INTRODUCTION

Within populations, individuals show repeatable and heritable differences in many behavioral traits (i.e., "animal personality") (Bell et al. 2009; Dochtermann et al. 2015, 2019). Individual differences in behaviors facilitating resource acquisition at the cost of survival, such as anti-predator behavior, aggressiveness, or exploration behavior (Stamps 2007; Wolf et al. 2007), may, for example, be maintained as an adaptation to variation in intrinsic or extrinsic conditions (Dingemanse and Wolf 2010; Wolf and Weissing 2010, 2012; Sih et al. 2015). Differences may also reflect alternative solutions to resolving life-history trade-offs (Stamps 2007; Wolf et al. 2007; Smith and Blumstein 2008), as suggested by the existence of relationships between behavior, physiology and life-history traits, called pace-of-life syndromes, found in certain populations or species (Réale et al. 2010; Dammhahn et al. 2018; Moiron et al. 2019).

Empirically testing adaptive "personality theory" ideally requires field-based behavioral assays (Archard and Braithwaite 2010; Niemelä and Dingemanse 2014). Those are difficult to acquire because of logistical constraints preventing the collection of repeated measures for suites of behaviors (Niemelä and Dingemanse 2018a), the necessity to separate individual effects from environmental confounds (Martin and Réale 2008; Westneat et al. 2011; Niemelä and Dingemanse 2017) and the necessity of large sample sizes required for parameter estimation (Martin et al. 2011; van de Pol 2012; Dingemanse and Dochtermann 2013; Niemelä and Dingemanse 2018a). Behavior is therefore often measured in controlled laboratory conditions, such that large numbers of wild-caught individuals can be typed, and subsequently linked to fitness in the wild (Dingemanse and Réale 2005; Réale et al. 2007; Smith and Blumstein 2008; Moiron et al. 2020).

Meta-analyses, however, have demonstrated misfits between adaptive theory and empirical data (Garamszegi et al. 2012; Niemelä and Dingemanse 2018b; Royauté et al. 2018). Relying on laboratory-based tests assumes the assayed behavior predicts behavior in the wild. Gene-by-environment interactions can, however, change the expression of genetic variation (Hoffmann and Merilä 1999; Charmantier and Garant 2005; Hodgins-Davis and Townsend 2009; Dochtermann et al. 2019). That is, rank orders among individuals might change (Dingemanse, Barber, et al. 2020), such that individuals that are, for example, relatively active in the field are not also relatively active in the laboratory (Herborn et al. 2010; Niemelä and Dingemanse 2014). Rank-order changes are likely for behavior because this type of trait is quickly adjusted to the environment (Dingemanse et al. 2009, 2020; Hodgins-Davis and Townsend 2009), such as laboratory versus field conditions (Archard and Braithwaite 2010;

Niemelä and Dingemanse 2014). Thus, laboratory-based assays must be validated in the field to ensure both are expressions of the same latent trait (or "character") (Houle et al. 2011; Carter et al. 2013; Araya-Ajoy and Dingemanse 2014). Across a range of taxa, cross-validation has been performed (Herborn et al. 2010; Cole and Quinn 2014; Fisher et al. 2015; McCowan et al. 2015; Yuen et al. 2016; Osborn and Briffa 2017; Edwards et al. 2018). In many of these cases, different functional types of responses were used, such as activity in a novel environment in the laboratory vs. discovery of feeders in the wild. Such validation may measure syndromes between functionally distinct behaviors rather than the same underlying latent character (but see Yuen et al. 2016). Validation is often also achieved indirectly, by demonstrating that both laboratory- and field-based behavior assayed in one functional context predict field-based behavior in another. For example, in great tits (Parus major), field-based measures of behaviors related to risk-taking (e.g. aggressiveness toward a conspecific or human intruder, boldness toward a novel object, anti-predator behavior) correlate with 'exploration behavior' scored as the number of movements in a novel environment assayed in a laboratory room (Verbeek et al. 1994; Hollander et al. 2008; Quinn et al. 2009; Amy et al. 2010; Stuber et al. 2013), but also with the number of movements scored in a cage in the field (Mutzel et al. 2013; Stuber et al. 2013). Such studies typically suggest that both latter behaviors represent alternative proxies for risk-taking in the wild (Dingemanse and De Goede 2004; Hollander et al. 2008; Stuber et al. 2013; Abbey-Lee and Dingemanse 2019; Moiron et al. 2019). A key question, however, is whether a direct validation of those behaviors scored in the laboratory versus the field would show cross-context consistency.

Specifically, given that in great tits laboratory- and field-based activity in a novel environment both correlate with similar types of field-based behaviors related to risk-taking, they might both be expressions of a single latent character (Fig. 1A). However, they might also be expressions of distinct latent characters that nevertheless each affect similar field-based behaviors (Fig. 1B). In the first scenario, we expect positive correlations between laboratory- and field-based behavioral scores (Araya-Ajoy and Dingemanse 2014; Dingemanse, Barber, et al. 2020), while in the second scenario, we expect no correlation. Here, we tested the key assumption that activity scores in both novel environments correlate positively, thus representing expressions of the same latent character (Fig. 1A).

While testing this hypothesis, we considered that phenotypic correlations can be misleading because they comprise influences of (co)variances occurring at multiple hierarchical levels (Searle 1961). For example, positive among-individual correlations between laboratory- and field-based traits might not be visible in phenotypic data if the two traits are

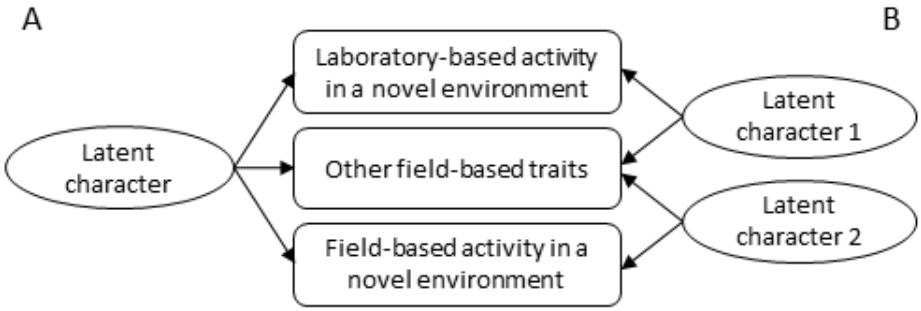


Fig. 1. Two scenarios for relationships between laboratory- and field-based behavior. In scenario A, laboratory- and field-based behavior correlate with each other and other field-based behaviors, all representing expressions of a single latent character. In scenario B, laboratory- and field-based behavior do not correlate with each other but with other field-based behaviors, thus representing expressions of two distinct latent characters (1 and 2). This study on great tits showed that the activity in a novel environment assayed in the laboratory and in the field did not correlate, thus, supported scenario B.

weakly repeatable or heritable, while measurement error or other sources of within-individual variation are simultaneously uncorrelated across contexts (Dingemanse, Dochtermann, et al. 2012; Brommer 2013; Dingemanse and Dochtermann 2013). Similarly, genetic correlations might be hidden at the phenotypic level if permanent environmental correlations are opposite in sign (van Noordwijk and de Jong 1986; Hadfield et al. 2007; Santostefano et al. 2017). We thus applied quantitative genetics approaches (Kruuk 2004; Wilson et al. 2010) to partition the variation in laboratory- and field-based behavior in among- and within-individual components and estimate repeatability. We then partitioned the among-individual variance into additive genetic and permanent environmental effects to estimate heritability (Wilson et al. 2010). These analyses were warranted because laboratory-based behavior is heritable (range h^2=0.10-0.30) based on data collected from various wild great tit populations (Dingemanse et al. 2002, 2004; Quinn et al. 2009; Nicolaus et al. 2012, 2013; Korsten et al. 2013). We also estimated among- and within-individual correlations, and partitioned the former into additive genetic versus permanent environmental correlations (Dingemanse and Dochtermann 2013).

Our analyses show that activity scores do not correlate across laboratory- and field-based novel environment assays. This absence of correlation could be genuine but also result from attenuation effects of sampling bias on trait correlations (Neale et al. 1989; Carter et al. 2012). Sampling bias is a concern documented in animal personality research (Biro and Dingemanse 2009; Garamszegi et al. 2009; Carter et al. 2012), and has been detected in our species (Stuber et al. 2013). As a post hoc analysis, we therefore also assessed whether birds sampled in winter for laboratory-based tests represented a biased sample with respect to field-based phenotypes.

MATERIAL AND METHODS

Study sites and field procedures

Data were collected over a 10-year period (2010-2019) from 12 nest box plots south of Munich, Germany (47° 97'N, 11° 21'E), each fitted with 50 nest boxes. Data collection consisted of roosting inspections in winter and monitoring of the breeding population in spring/summer. All procedures complied with the guidelines from the District Government of Upper Bavaria (Regierung von Oberbayern) for Animal care, permit no. 55.2-1-54-2532-140-11 and ROB-55.2-2532.Vet_02-17-215.

Once per winter, we checked all nest boxes at night for roosting birds within a two-week period (in January-February, and additionally November-December in 2010-2011 only). Each roosting bird was ringed if previously un-ringed, weighed, placed in a carrying box, and transported to the laboratory. Upon arrival, birds were again weighed, sexed (based on plumage characteristics; Jenni and Winkler 1994) and placed alone in a standard cage with water and food (sunflower seeds) ad libitum. The next morning (08h00-13h00), birds were individually released into a novel environment via a sliding door, such that they did not have to be handled. The novel environment was a standard (5.2L × 2.9W × 2.3H m3) laboratory room furbished with five artificial trees. Each tree consisted of four 20 cm long branches, fitted on a 4L × 4W × 150H cm³ trunk (Dingemanse et al. 2002). One minute before releasing the bird, the observer placed a towel on the front grid of the cage, and gently lifted it after opening the sliding door. This stimulated birds to enter the room without physical handling. We calculated an activity score in this novel environment by totaling the number of flights and hops between perches made during the first two minutes after the bird entered the room (Dingemanse et al. 2002). All birds were released at their capture location within 24h of capture. We conducted a total of 1377 tests on 1011 individuals, with an average of 153.00 (standard deviation (sd) = 59.80) birds assayed per year.

During the breeding season (April-July), nest boxes were checked at least once per week to record breeding parameters (Nicolaus et al. 2015; Dingemanse, Moiron, et al. 2020). Breeding birds were caught in their nest box using a spring trap when their nestlings were 7-12 days old. Birds were ringed if previously unbanded, and immediately placed into a holding box (11L × 12W × 11H cm³) connected to a cage (61L × 39W × 40H cm³). Prior to capture, the cage was positioned at a distance of at least 50m from the focal nest box. It consisted of nontransparent material, but the front was fitted with a metal grid. Three perches were placed at fixed positions (illustrated in Stuber et al. 2013). The holding box was darkened with a bird

bag to help the bird to settle down. After one minute, the observer opened the sliding door of the holding box and removed the bird bag, thereby stimulating the bird to jump into the cage without handling, which was effective in all cases. We video-recorded its behavior for the first two minutes following cage entry. The setup was designed to represent a transportable version of the laboratory-based novel environment test. As above, an activity score was extracted from the recorded data, calculated as the number of movements made between three floor sections, six sections on the grid and the three perches (illustrated in Stuber et al. 2013). All birds were tested between 07:00 and 16:00. After the test, birds were measured for morphology and then released near their nest box. We conducted a total of 3648 tests on 2326 individuals, with an average of 364.80 (sd = 59.53) birds assayed per year.

Statistical analyses

We first fitted univariate mixed-effects (animal) models to separately estimate the sources of variation in laboratory- and field-based behavior. We then used bivariate mixed-effects (animal) models to estimate the cross-context correlation between laboratory- and field-based behavior. We ran all models in ASReml-R 4.0 (Butler et al. 2017).

We used all observations of all individuals scored in either one or both types of tests. Our dataset comprised 2647 individuals and 4305 observations; 657 records had only a laboratory score, 2928 only a field score, and 720 both scores. A minority of individuals was scored twice within a season (laboratory score: 72 out of 1011 individuals, 7.1%; field score: 452 out of 2326 individuals, 19.4%). We pragmatically decided to use only the first score of an individual in a focal season so that all individuals were treated equally. The average bird was assayed for laboratory- vs. field-based behavior 1.36 vs. 1.57 times. The mean activity score was 18.88 (sd = 11.78) for the laboratory and 65.13 (sd = 21.79) for the field test. Our social pedigree was based on 335 individuals for which the identity of both parents was known, and 15 and 6 individuals for which, respectively, only mother's or father's identity was known. Mothers from 300 (of 350 for which the mother was known) and fathers from 201 (of 341 for which the father was known) individuals were scored for laboratory-based behavior. All parents but two mothers were scored for field-based behavior. Ninety-five individuals had one to three siblings and 73 individuals had one of their grand-parents scored for at least one behavior.

Univariate mixed-effects animal models

We fitted random-intercept mixed-effects (animal) models for each type of assay separately. These models included various fixed effects aiming at controlling for biases caused by the

experimental design. We fitted sex (males coded as -0.5, females as 0.5, fitted as covariate; Gelman 2008) and sequence (i.e., test number within individual within context (laboratory or field); range:1-6 tests for the laboratory and 1-7 for the field assay; fitted as fixed-effect factor). The latter was fitted to estimate within-individual variation in behavior beyond any habituation effect. We examined whether the presence of sequence categories with few data points biased our estimates (Table S1 and S2), which was not the case (Table S3). We also fitted the covariate interval (number of days between two consecutive tests within context, with the first test being given the value zero) because previous studies showed that sequence effects are distinct from interval effects as memory of previous experiences appear to fade when inter-test intervals are long (Dingemanse et al. 2002). We further fitted time of day (time since sunrise in hours, mean-centered within context; fitted as covariate) to control for any diurnal changes. Finally, we also controlled for potential effects of various experiments (brood size manipulation, perceive predation risk manipulation and supplementary feeding; fully detailed in Appendix S1) performed in this population over the years. We fitted random intercepts for individual (VI, representing the summed total of additive genetic and permanent environmental effects, see below), test date, year, plot and plot-year (the unique combination of plot and year). The latter three factors controlled for, respectively, temporal, spatial, and spatiotemporal environmental variation (Araya-Ajoy et al. 2016). Date captured any (non)linear temporal effects within years, which thereby controlled for any unmeasured environmental effects changing with date (Dingemanse, Bouwman, et al. 2012).

All models used the inverse of the relatedness matrix inferred from the social pedigree to estimate the among-individual variance (VI) in behavior attributable to additive genetic (VA) versus permanent environmental effects (VPE) (Kruuk 2004; Wilson et al. 2010). For each type of score, we calculated repeatability (R) for each variance component as the proportion of total phenotypic variance (VP) explained by a focal variance component (e.g., repeatability among individuals $R=VI/VP=(VA+VPE)/VP$) (Nakagawa and Schielzeth 2010; Dingemanse and Dochtermann 2013), where VP represents the summed total of all variance components. We also calculated the narrow-sense heritability (h^2) as the proportion of total phenotypic variance (VP) explained by additive genetic effects ($h^2=VA/VP$) (Lynch and Walsh 1998; Nakagawa and Schielzeth 2010). Pedigree relationships were based on a social pedigree. Paternity misassignments stemming from extra-pair paternities can result in misestimated heritability estimates. However, extra-pair paternity rate in our population is relatively low (9.6%; Araya-Ajoy et al. 2016). Simulation studies show, moreover, that heritability estimates based on social

pedigrees are typically unbiased with data structures such as ours (Charmantier and Réale 2005).

Statistical significance of variance components was assessed using likelihood ratio tests (LRTs), a statistic following a Chi-square (χ^2) distribution, estimated as twice the deviance in loglikelihood between a focal full model and the same model excluding a focal random effect. P-values were calculated assuming an equal mixture of χ^2 (df=0) and χ^2 (df=1) because variances are zero-positive bounded (Miller 1977; Stram and Lee 1994). Statistical significance of fixed effects was based on conditional Wald F-tests (Butler et al. 2017).

Bivariate mixed-effects animal model

We fitted a bivariate mixed-effects (animal) model, with laboratory- and field-based behavior as the two response variables, and estimated their correlation among individuals, within individuals, among plots and among years. The fixed and random effect structure was identical to the univariate models. For date, the cross-context covariance was not estimable by design, and thus constrained to zero (Dingemanse and Dochtermann 2013). Because the plot-year variance was negligible for field-based behavior (Table S4), the covariance between contexts for plot-year caused model convergence failure. We thus constrained the plot-year variance in field-based behavior to zero, as well as the covariance at this level.

Statistical significance of covariances was assessed using LRTs based on the deviance between the full model and a model where a focal covariance was fixed to zero, and where the p-value was calculated assuming one degree of freedom.

Sampling bias

To test whether the lack of correlation between field- and laboratory-based scores (see Results) resulted from sampling bias, we performed an *a posteriori* conceived analysis. To the univariate model for field-based behavior, we added a 2-level factor that distinguished breeders that had been caught, vs. not caught, roosting in the preceding winter. To acknowledge that a difference in behavior might be due to both among-individual (i.e., selective sampling or appearance) and within-individual effect (i.e., phenotypic plasticity), as a second step, we disentangled these two effects by applying the within-subject centering approaches (Appendix S2). For further explanation of the utility of this within-subject centering approach, see van de Pol and Wright (2009).

RESULTS

Laboratory-based behavior

Laboratory-based activity in a novel environment varied among individuals (V_I±SE=44.29±5.28, p<0.001; R=0.35±0.04; Table S4). The among-individual variance was attributable to permanent environmental (V_{PE}=35.57±12.90, p=0.001; pe^2=0.30±0.11) rather than additive genetic effects (V_A=6.75±12.25, p=0.26; h^2=0.06±0.10). Laboratory-based behavior was significantly repeatable but not significantly heritable. This finding implies a lack of evidence for additive genetic effects that would cause resemblance between related individuals. Laboratory-based behavior further varied among years and plots, but not among plot-years or test dates (Table S4). Finally, individuals moved more in repeat tests (sequence effect, Fig. 2a, Table S1), but the sequence effect essentially weakened with increasing inter-test interval (β±SE=-0.01±0.003, p=0.01). See supplementary materials for effect sizes of other fixed effects (Table S1).

Field-based behavior

Field-based activity in the novel environment also varied among individuals (V_I=164.05±11.72, p<0.001; R=0.37±0.02; Table S4). The among-individual variance was attributable to both permanent environmental (V_{PE}=62.66±24.37, p=0.002; pe^2=0.14±0.05) and additive genetic effects (V_A=101.60±24.49, p<0.001; h^2=0.23±0.05). Field-based behavior was thus both significantly repeatable and significantly heritable. This finding implies evidence for additive genetic effects causing resemblance between related individuals. Field-based behavior further varied among years, plots and test dates, but not among plot-years (Table S4). Finally, individuals moved less (rather than more, see above) over repeat tests (sequence effect, Fig.2b, Table S2), and moved more when tested later in the day (time of day effect; β=0.50±0.17, p=0.003; Table S2). See supplementary materials for effect sizes of other fixed effects (Table S2).

Cross-context correlations

Laboratory- and field-based behaviors did not correlate at any hierarchical level (Fig.3, Table 1). That is, they correlated neither among (r_{among}±SE=0.02±0.07, p=0.77) nor within individuals (r_{within}=0.01±0.05, p=0.85), and neither temporally (r_{year}=0.04±0.59, p=0.95) nor spatially (r_{plot}=0.54±0.35, p=0.19). Decomposition of the among-individual correlation revealed that neither the additive genetic (r_G=-0.44±0.84, p=0.45) nor the permanent

environmental (r_{PE}=0.23±0.27, p=0.37) correlations differed from zero. We note that the genetic correlation estimate was highly uncertain, which often occurs when the additive genetic variance in one trait (here laboratory-based behavior) is close to zero, and therefore hard to estimate precisely (Roff 2001). Nevertheless, the genetic variance was greater in field-based behavior, which indicates a difference in additive genetic variance in behavior between the two contexts. The absence of a correlation overall implies that laboratory- and field-based behaviors are genetically independent (Roff 1997) and belong to two distinct latent characters (Fig.1B).

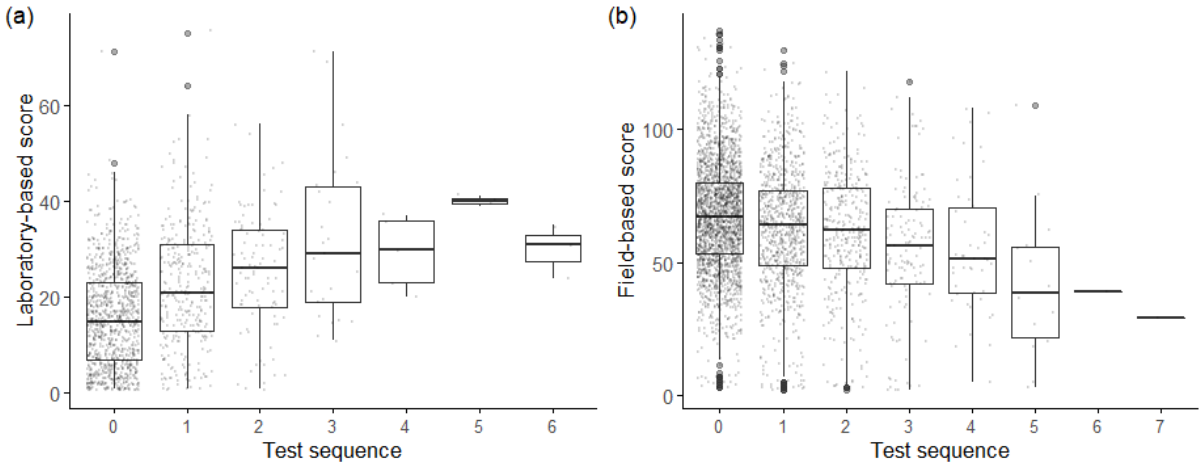


Fig. 2. Boxplots, with whiskers, show the median, first and third quartile of sequence effects for (a) laboratory- and (b) field-based behavior, and small dots the scores of each individual within a focal sequence. Large dots above or below whiskers are outliers. Laboratory-based score increased over sequence, whereas field-based score decreased.

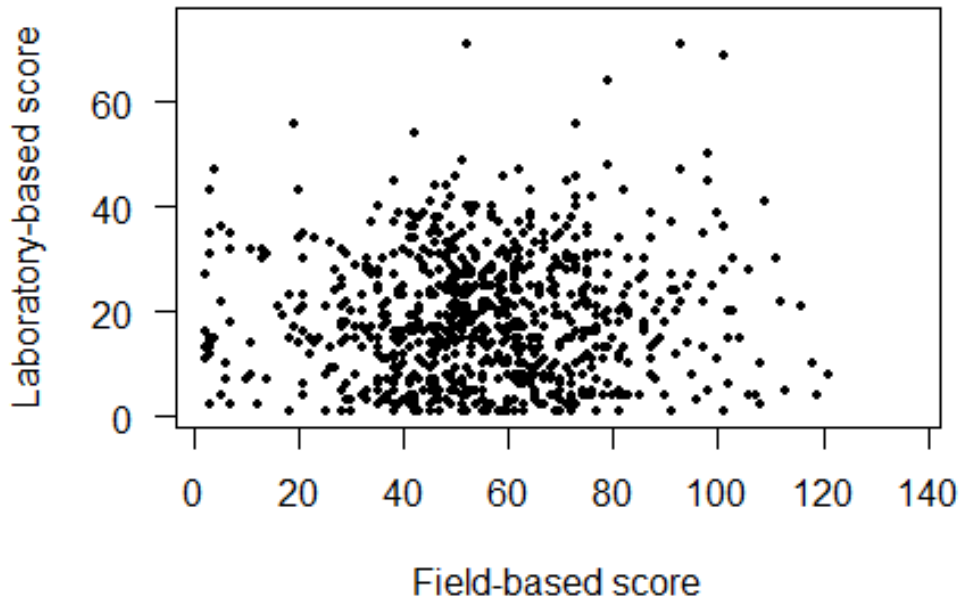


Fig.3. Absence of correlation between laboratory- and field-based behavior. Each dot represents an observation of an individual scored for both laboratory- and field-based behavior within the same year.

Table 1. Correlations between laboratory- and field-based behavior. Correlation ($r_{lab\text{-}field}$) estimates are presented with their standard errors (se) and are derived from a bivariate mixed-effects model. The random effect plot-year was fitted, but the correlation not estimated because of low plot-year variance in field-based behavior causing model convergence failure (see Methods). Statistical significance (p-value) was tested using log-likelihood ratio tests, from which χ^2-values were derived and considering one degree of freedom (χ^2_1).

Correlations	$r_{lab\text{-}field}$ (se)	χ^2_1	p-value
Among year	0.04 (0.59)	0.36	0.82
Among plot	0.54 (0.35)	2.58	0.11
Among individual	0.02 (0.07)	0.09	0.76
Additive genetic	-0.44 (0.84)	1.34	0.25
Permanent environmental	0.23 (0.27)	1.55	0.21
Within individual (residual)	0.010 (0.05)	0.85	0.36
Total phenotypic	0.03 (0.04)	0.91	0.34

Sampling bias

Birds captured as breeders that we also caught roosting in the preceding winter were, on average, 15% less active in the field-based test than the ones not caught in winter. This decrease in behavior was due to both among- and within-individual effects, i.e., selective sampling and phenotypic plasticity (Appendix S2, Fig.S1, Table S5). The population sampled in winter was thus biased toward slower explorers (an among-individual effect) with respect to field-based behavior. Follow-up data simulations showed this sampling bias was insufficient to bias the cross-context correlation (Table S6). Interestingly, this analysis also revealed that an individual that was caught in the winter preceding a focal breeding season explored the cage 12% less than when it was not caught (a within-individual effect). Winter captures thus came with carry-over effects on field-based behavior scored in spring. The effects of selective sampling and phenotypic plasticity were statistically distinct (Table S5; see Appendix S2 for details), implying they resulted from two distinct biological processes (van de Pol and Wright 2009).

DISCUSSION

Tests of adaptive personality theory are often derived from laboratory-based behaviors assumed to predict behavior in the wild. Meta-analyses, however, report misfits between adaptive personality theory and empirical data (Niemelä and Dingemanse 2018b; Royauté et al. 2018). Misfits could result from laboratory-based scores not predicting field-based behavior (e.g. Carter et al. 2013). The importance of cross-validating this assumption is emphasized by our study showing no association between laboratory- and field-based assays of activity in a novel environment. This finding was unexpected because laboratory- and field-based scores

both predict a behavioral response toward a novel object (nest box camera) in our great tit population (Stuber et al. 2013). The absence of a cross-context correlation implies they reflect expressions of distinct latent characters under different selection pressures.

Laboratory- and field-based behaviors were both repeatable. However, the absence of a cross-context correlation implies that repeatable rank-order differences among individuals changed across contexts, suggesting strong gene-by-environment interactions (Niemelä and Dingemanse 2014). The absence of a correlation could be due to effects of season (non-breeding vs. breeding), time in captivity prior to testing (overnight vs. few seconds) or testing apparatus (room vs. cage). We exclude confounding effects of season because, within four independent great tit populations, individual rank order did not change from September to June based on the laboratory assay (Dingemanse, Bouwman, et al. 2012). Thus, laboratory-based behavior can be viewed as the same trait in different seasons. Rank-order differences among birds in laboratory-based scores should thus arguably not have changed if the assay had been conducted in the breeding season. We therefore expect the same cross-context correlation as all assays had been conducted in the same season. We appreciate that trait correlations may also change when only one of the two traits shows gene-by-environment interactions (for illustration, see Dingemanse et al. 2020; Mitchell and Houslay 2021). We assume, however, that field-based behavior also represents the same trait in different seasons; this assumption requires future testing. We also exclude effects of time prior to testing, which could induce individual variation in response to stress in the laboratory. Repeatable differences in laboratory-based behavior may represent cryptic genetic variation not normally expressed in the wild (Schlichting 2008). Cryptic variation might not be expressed in the field test because handling times prior to testing were short. If this assumption holds, the additive genetic variance should be greater for the laboratory versus the field test (Schlichting 2008). However, we observed the opposite pattern, making this explanation unlikely. Of course, field-based behavior could also represent an immediate stress response induced by capture, implying it does not reflect genuine activity in a novel environment. However, the stress response in great tits has previously been shown to correlate with laboratory-based behavior (assayed in very similar conditions), with birds moving less also responding faster to stressors (Baugh et al. 2013). Individual variation in stress response and activity in the laboratory-based novel environment therefore appear to match, rendering bias caused by differences in time prior to testing arguably unlikely. Testing apparatus, by contrast, could explain the lack of cross-context correlation for two reasons.

First, birds may perceive both apparatuses differently due to size differences and thus express different behaviors. Indeed, great tits appear to express different behaviors when the

size and the complexity of the laboratory room is experimentally manipulated (Arvidsson et al. 2017). Second, in the laboratory, birds were largely isolated from external stimuli (though they were able to hear birds in adjacent cages), whereas in the field, birds could perceive (hear and see) most components of the external environment through the grid. This difference in perception might affect their behavior because birds would likely recognize their surroundings when assayed in the field and because the presence/absence of social cues can induce changes in behavioral expression (Rudin et al. 2018, 2019). Indeed, studies on Australian field crickets (Teleogryllus oceanicus) also found that the genetic covariances of behaviors related to boldness, exploration and activity differed between acoustic and silent environments (Rudin et al. 2019). One plausible functional interpretation is that this difference in perception may cause birds to express a combination of anxiety and exploration of the novel environment in the laboratory-based test, but an escape behavior to return to their territory in the field-based test. This interpretation is fully consistent with our finding that activity increased with test sequence in the laboratory (room) but instead decreased in the field (cage). If birds behave less spatially neophobic and more neophilic in the laboratory, they should express more motivation in exploring with repeated exposure (Mettke-Hofmann et al. 2009; Greggor et al. 2015). This motivation appeared to fade as the interval between tests increased, that is, as memory effects vanished. In birds, reduction in movements or freezing may compare to thigmotaxis (i.e., time spent close to walls) and freezing in rodents in an open-field test. Indeed, these latter behaviors appeared to predict anxiety rather than exploration as they did not correlate with other movements recorded in the open-field (Choleris et al. 2001; Krebs et al. 2019). By contrast, birds might be less motived to escape, when familiarity with the environment (cage) increases with repeated exposure. Consistent with this explanation, birds spent most of their time pecking the grid (pers. obs. of both authors). The motivation to escape appeared to be modulated by the time of day, which may relate to food demand of their nestlings, or diurnal variation in predation risk. An experimental test would require redesigning the field-based test such that birds do not have any contact with the external environment (Charmantier et al. 2017; Dubuc-Messier et al. 2017). We note, however, that the field-based test was specifically designed to assess behavioral variation in a natural environment, where social cues or any other environmental factors may influence behavioral expression (van Overveld and Matthysen 2013; Rudin et al. 2018; Smit and van Oers 2019). Assays disentangling anxiety from exploration would be required to better understand the function of laboratory- and field-based activity in a novel environment (Mettke-Hofmann et al. 2009; Greggor et al. 2015).

The lower additive genetic variance in laboratory-based behavior might reflect ecological interactions, resulting in biased sampling if certain behavioral types do not roost in boxes or are immigrants arriving after the winter sampling. Biased sampling can attenuate estimates of cross-context correlations (Neale et al. 1989; Carter et al. 2012). Indeed, the winter sample was biased toward relatively less active individuals, which may be due to less active birds preferably roosting in nest boxes or to immigrants arriving late winter being relatively more active than residents. This latter explanation is likely as more active, explorative or bold individuals generally have been shown to be more dispersive or migratory (Cote et al. 2010; Korsten et al. 2013; van Overveld et al. 2014). The magnitude of the bias detected in our study was, however, insufficient to substantially affect the cross-context correlation.

The absence of a correlation between two seemingly similar behaviors calls for major caution in functional labeling. Our study implies that care is required in labeling behaviors seemingly assaying similar functions but scored using dissimilar methods (e.g., size or complexity of apparatuses, presence/absence of social cues) prior to firm validation (Carter et al. 2013; Arvidsson et al. 2017). Researchers commonly compare their own results to studies applying the same label rather than distinctly test different paradigms. We ourselves previously used findings from other great tit populations linking fitness and laboratory-based behavior (e.g., Dingemanse et al. 2004; Both et al. 2005; Serrano-Davies et al. 2017) to predict links between fitness and field-based behavior (Nicolaus et al. 2015; Araya-Ajoy et al. 2016; Abbey-Lee and Dingemanse 2019). This study, however, implies that the laboratory- and field-based tests used here measure distinct characters, each of which may have evolved in response to different selection pressures. Interestingly, the two types of scores have previously been shown to each positively and additively explain variation in behavioral response toward a novel object (Stuber et al. 2013). This agrees with our conclusions that they represent expressions of different latent characters, provided that both behaviors relate to risk or resource acquisition independently. Field-based behavior correlated positively with male aggressiveness during conspecific territorial intrusions in the wild in our population (Moiron et al. 2019), while laboratory-based behavior positively predicted nest defense behavior toward human intruders in a Belgian population (Hollander et al. 2008) and dominance among territorial birds at feeders in a Dutch population (Dingemanse and De Goede 2004). Repeatable differences in aggressiveness might thus be influenced by multiple underlying latent traits, such as exploration and escape behavior. Alternatively, aggressiveness assayed in different contexts may also not correlate, and thus also stem from different underlying characters.

Repeatability did not differ between laboratory- and field-based behavior but heritability did. The application of a quantitative genetics approach thus revealed that despite exhibiting similar phenotypic variances, these two traits may differ substantially in evolutionary potential. This finding also implies they measure distinct characters, corroborating our finding of a lack of cross-context correlation. Laboratory-based behavior was evidently affected mostly by permanent environmental effects (e.g., parental effects, epigenetics or environmental conditions during development). We note that the heritability of laboratory-based behavior was lower than the heritability reported for other great tit populations, among which the estimates also somewhat differ (Westerheide, the Netherlands: h^2(se)=0.22±0.14 (Dingemanse et al. 2002); Lauwersmeer, the Netherlands: h^2=0.10±0.05 (Nicolaus et al. 2012); Wytham Woods, United Kingdom: h^2=0.23±0.07 (Quinn et al. 2009); Boshoek, Belgium: h^2=0.30±0.11 (Korsten et al. 2013)). A post-hoc analysis showed that our low estimate of heritability was not simply caused by our sampling design with fewer individuals tested in the laboratory vs. the field. That is, when we performed exactly the same analyses on another trait that was measured alongside activity (body mass), we found that its heritability did not differ between the laboratory (h^2=0.34±0.13) and the field (h^2=0.34±0.06). Laboratory-based behavior assayed and scored using very similar procedures may thus be under selection pressures that differ among populations and resulting from spatiotemporal variation in environmental conditions (Foster and Endler 1999; Siepielski et al. 2009; Siepielski et al. 2013). By contrast, additive genetic and permanent environmental effects appeared to affect field-based behavior to a more similar degree. Among-study replication is now required to test the generality of these differences by means of a comparative quantitative genetics study.

In conclusion, activity in laboratory- and field-based novel environments, both previously labeled “exploration behavior”, represent independent traits likely having distinct evolutionary trajectories given their genetic architecture. Laboratory- and field-based behaviors likely are expressions of multiple quasi-independent behavioral characters (Araya-Ajoy and Dingemanse 2014), underpinned by different proximate mechanisms and selective pressures. Seemingly similar response variables might thus represent entirely different aspects of an individual’s behavior when derived from different assays due to gene-by-environment interactions, as previously highlighted in other sub-fields of evolutionary ecology (Charmantier and Garant 2005; Hodgins-Davis and Townsend 2009). Assays may differ in many ways, including apparatus size, complexity or access to cues about the external environment. Gene-by-environment interactions likely also exist in other taxa where ‘exploration behavior’ is assayed

in open-field tests, arenas or aquaria, or for other repeatable and heritable behaviors, such as ‘boldness’ or ‘aggressiveness’.

ACKNOWLEDGMENTS

We thank past members of the Evolutionary Ecology of Variation group (Max Planck Institute for Ornithology, 2012-2017) and members, field assistants and students of the Behavioural Ecology group (LMU) for helping to collect the data. We thank Petri Niemelä and Raphaël Royauté for discussion. A.M. was part of the International Max-Planck Research School for Organismal Biology. This work was supported by the German Science Foundation (grant no. DI 1694/1-1) and by the Max-Planck Society (2012-2017).

Data accessibility: Analyses reported in this article can be reproduced using the data provided by Mouchet (2021)

REFERENCES

Abbey-Lee RN, Dingemanse NJ. 2019. Adaptive individual variation in phenological responses to perceived predation levels. Nat Commun. 10(1). doi:10.1038/s41467-019-09138-5. http://dx.doi.org/10.1038/s41467-019-09138-5.

Amy M, Sprau P, De Goede P, Naguib M. 2010. Effects of personality on territory defence in communication networks: A playback experiment with radio-tagged great tits. Proc R Soc B Biol Sci. 277(1700):3685–3692. doi:10.1098/rspb.2010.0598.

Araya-Ajoy YG, Dingemanse NJ. 2014. Characterizing behavioural "characters": An evolutionary framework. Proc R Soc B Biol Sci. 281(20132645):1–10. doi:10.1098/rspb.2013.2645.

Araya-Ajoy YG, Kuhn S, Mathot KJ, Mouchet A, Mutzel A, Nicolaus M, Wijmenga JJ, Kempenaers B, Dingemanse NJ. 2016. Sources of (co)variation in alternative siring routes available to male great tits (Parus major). Evolution (N Y). 70(10):2308–2321. doi:10.1111/evo.13024.

Archard GA, Braithwaite VA. 2010. The importance of wild populations in studies of animal temperament. J Zool. 281(3):149–160. doi:10.1111/j.1469-7998.2010.00714.x.

Arvidsson LK, Adriaensen F, van Dongen S, De Stobbeleere N, Matthysen E. 2017. Exploration behaviour in a different light: testing cross-context consistency of a common personality trait. Anim Behav. 123:151–158. doi:10.1016/j.anbehav.2016.09.005. http://dx.doi.org/10.1016/j.anbehav.2016.09.005.

Baugh AT, van Oers K, Naguib M, Hau M. 2013. Initial reactivity and magnitude of the acute stress response associated with personality in wild great tits (Parus major). Gen Comp Endocrinol. 189:96–104. doi:10.1016/j.ygcen.2013.04.030. http://dx.doi.org/10.1016/j.ygcen.2013.04.030.

Bell AM, Hankison SJ, Laskowski KL. 2009. The repeatability of behaviour: a meta-analysis. Anim Behav. 77(4):771–783. doi:10.1016/j.anbehav.2008.12.022. http://dx.doi.org/10.1016/j.anbehav.2008.12.022.

Biro PA, Dingemanse NJ. 2009. Sampling bias resulting from animal personality. Trends Ecol Evol. 24(2):63–66.

Both C, Dingemanse NJ, Drent PJ, Tinbergen JM. 2005. Pairs of extreme avian personalities have highest reproductive success. J Anim Ecol. 74(4):667–674. doi:10.1111/j.1365-2656.2005.00962.x.

Brommer JE. 2013. On between-individual and residual (co)variances in the study of animal personality: Are you willing to take the "individual gambit"? Behav Ecol Sociobiol. 67(6):1027–1032. doi:10.1007/s00265-013-1527-4.

Butler DG, Cullis BR, Gilmour AR, Gogel BJ, Thompson R. 2017. ASReml-R Reference Manual Version 4. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. http://www.homepages.ed.ac.uk/iwhite/asreml/uop.

Carter AJ, Feeney WE, Marshall HH, Cowlishaw G, Heinsohn R. 2013. Animal personality: What are behavioural ecologists measuring? Biol Rev. 88(2):465–475. doi:10.1111/brv.12007.

Carter AJ, Heinsohn R, Goldizen AW, Biro PA. 2012. Boldness, trappability and sampling bias in wild lizards. Anim Behav. 83(4):1051–1058. doi:10.1016/j.anbehav.2012.01.033. http://dx.doi.org/10.1016/j.anbehav.2012.01.033.

Charmantier A, Demeyrier V, Lambrechts M, Perret S, Grégoire A. 2017. Urbanization is associated with divergence in pace-of-life in great tits. Front Ecol Evol. 5(MAY):1–13. doi:10.3389/fevo.2017.00053.

Charmantier A, Garant D. 2005. Environmental quality and evolutionary potential: Lessons from wild populations. Proc R Soc B Biol Sci. 272(1571):1415–1425. doi:10.1098/rspb.2005.3117.

Charmantier A, Réale D. 2005. How do misassigned paternities affect the estimation of heritability in the wild? Mol Ecol. 14(9):2839–2850. doi:10.1111/j.1365-294X.2005.02619.x.

Choleris E, Thomas AW, Kavaliers M, Prato FS. 2001. A detailed ethological analysis of the mouse open field test: Effects of diazepam, chlordiazepoxide and an extremely low frequency pulsed magnetic field. Neurosci Biobehav Rev. 25(3):235–260. doi:10.1016/S0149-7634(01)00011-2.

Cole EF, Quinn JL. 2014. Shy birds play it safe: Personality in captivity predicts risk responsiveness during reproduction in the wild. Biol Lett. 10(5):8–11. doi:10.1098/rsbl.2014.0178.

Cote J, Clobert J, Brodin T, Fogarty S, Sih A. 2010. Personality-dependent dispersal: Characterization, ontogeny and consequences for spatially structured populations. Philos Trans R Soc B Biol Sci. 365(1560):4065–4076. doi:10.1098/rstb.2010.0176.

Dammhahn M, Dingemanse NJ, Niemelä PT, Réale D. 2018. Pace-of-life syndromes: a framework for the adaptive integration of behaviour, physiology and life history. Behav Ecol Sociobiol. 72(3):1–8. doi:10.1007/s00265-018-2473-y.

Dingemanse NJ, Barber I, Dochtermann NA. 2020. Non-consumptive effects of predation: does perceived risk strengthen the genetic integration of behaviour and morphology in stickleback? Ecol Lett. 23(1):107–118. doi:10.1111/ele.13413.

Dingemanse NJ, Both C, Drent PJ, Van Oers K, Van Noordwijk AJ. 2002. Repeatability and heritability of exploratory behaviour in great tits from the wild. Anim Behav. 64(6):929–938. doi:10.1006/anbe.2002.2006.

Dingemanse NJ, Both C, Drent PJ, Tinbergen JM. 2004. Fitness consequences of avian personalities in a fluctuating environment. Proc R Soc B Biol Sci. 271(1541):847–852. doi:10.1098/rspb.2004.2680.

Dingemanse NJ, Bouwman KM, van de Pol M, van Overveld T, Patrick SC, Matthysen E, Quinn JL. 2012. Variation in personality and behavioural plasticity across four populations of the great tit Parus major. J Anim Ecol. 81(1):116–126. doi:10.1111/j.1365-2656.2011.01877.x.

Dingemanse NJ, Dochtermann NA. 2013. Quantifying individual variation in behaviour: Mixed-effect modelling approaches. J Anim Ecol. 82(1):39–54. doi:10.1111/1365-2656.12013.

Dingemanse NJ, Dochtermann NA, Nakagawa S. 2012. Defining behavioural syndromes and the role of "syndrome deviation" in understanding their evolution. Behav Ecol Sociobiol. 66(11):1543–1548. doi:10.1007/s00265-012-1416-2.

Dingemanse NJ, De Goede P. 2004. The relation between dominance and exploratory behavior is context-dependent in wild great tits. Behav Ecol. 15(6):1023–1030. doi:10.1093/beheco/arh115.

Dingemanse NJ, Moiron M, Araya-Ajoy YG, Mouchet A, Abbey-Lee RN. 2020. Individual variation in age-dependent reproduction: Fast explorers live fast but senesce young? J Anim Ecol. 89(2):601–613. doi:10.1111/1365-2656.13122.

Dingemanse NJ, Van Der Plas F, Wright J, Réale D, Schrama M, Roff DA, Van Der Zee E, Barber I. 2009. Individual experience and evolutionary history of predation affect expression of heritable variation in fish personality and morphology. Proc R Soc B Biol Sci. 276(1660):1285–1293. doi:10.1098/rspb.2008.1555.

Dingemanse NJ, Réale D. 2005. Natural selection and animal personality. Behaviour. 142(9):1159–1184. doi:10.1163/156853905774539445.

Dingemanse NJ, Wolf M. 2010. Recent models for adaptive personality differences: A review. Philos Trans R Soc B Biol Sci. 365(1560):3947–3958. doi:10.1098/rstb.2010.0221.

Dochtermann NA, Schwab T, Anderson Berdal M, Dalos J, Royauté R. 2019. The Heritability of Behavior: A Meta-analysis. J Hered. 110(4):403–410. doi:10.1093/jhered/esz023.

Dochtermann NA, Schwab T, Sih A. 2015. The contribution of additive genetic variation to personality variation: Heritability of personality. Proc R Soc B Biol Sci. 282(1798):1–5. doi:10.1098/rspb.2014.2201.

Dubuc-Messier G, Reále D, Perret P, Charmantier A. 2017. Environmental heterogeneity and population differences in blue tits personality traits. Behav Ecol. 28(2):448–459. doi:10.1093/beheco/arw148.

Edwards DD, Rapin KE, Moore PA. 2018. Linking phenotypic correlations from a diverse set of laboratory tests to field behaviors in the crayfish, Orconectes virilis. Ethology. 124(5):311–330. doi:10.1111/eth.12734.

Fisher DN, James A, Rodríguez-Muñoz R, Tregenza T. 2015. Behaviour in captivity predicts some aspects of natural behaviour, but not others, in a wild cricket population. Proc R Soc B Biol Sci. 282(1809). doi:10.1098/rspb.2015.0708.

Foster SA, Endler JA. 1999. Geographic variation in behavior. Perspectives on evolutionary mechanisms. Foster SA, Endler JA, editors. Oxford University Press.

Garamszegi LZ, Eens M, Török J. 2009. Behavioural syndromes and trappability in free-living collared flycatchers, Ficedula albicollis. Anim Behav. 77(4):803–812. doi:10.1016/j.anbehav.2008.12.012.

Garamszegi LZ, Markó G, Herczeg G. 2012. A meta-analysis of correlated behaviours with implications for behavioural syndromes: Mean effect size, publication bias, phylogenetic effects and the role of mediator variables. Evol Ecol. 26(5):1213–1235. doi:10.1007/s10682-012-9589-8.

Gelman A. 2008. Scaling regression inputs by dividing by two standard deviations. Stat Med. 27:2865–2873. doi:10.1002/sim.

Greggor AL, Thornton A, Clayton NS. 2015. Neophobia is not only avoidance: Improving neophobia tests by combining cognition and ecology. Curr Opin Behav Sci. 6:82–89. doi:10.1016/j.cobeha.2015.10.007. http://dx.doi.org/10.1016/j.cobeha.2015.10.007.

Hadfield JD, Nutall A, Osorio D, Owens IPF. 2007. Testing the phenotypic gambit: Phenotypic, genetic and environmental correlations of colour. J Evol Biol. 20(2):549–557. doi:10.1111/j.1420-9101.2006.01262.x.

Herborn KA, Macleod R, Miles WTS, Schofield ANB, Alexander L, Arnold KE. 2010. Personality in captivity reflects personality in the wild. Anim Behav. 79(4):835–843. doi:10.1016/j.anbehav.2009.12.026. http://dx.doi.org/10.1016/j.anbehav.2009.12.026.

Hodgins-Davis A, Townsend JP. 2009. Evolving gene expression: from G to E to G × E. Trends Ecol Evol. 24(12):649–658. doi:10.1016/j.tree.2009.06.011.

Hoffmann AA, Merilä J. 1999. Heritable variation and evolution under favourable and unfavourable conditions. Trends Ecol Evol. 14(3):96–101.

Hollander FA, Van Overveld T, Tokka I, Matthysen E. 2008. Personality and nest defence in the great tit (Parus major). Ethology. 114(4):405–412. doi:10.1111/j.1439-0310.2008.01488.x.

Houle D, Pélabon C, Wagner GP, Hansen TF. 2011. Measurement and meaning in biology. Q Rev Biol. 86(1):3–34.

Jenni L, Winkler R. 1994. Moult and ageing of european passerines. First edit. London, UK: Academic Press.

Korsten P, Van Overveld T, Adriaensen F, Matthysen E. 2013. Genetic integration of local dispersal and exploratory behaviour in a wild bird. Nat Commun. 4(May 2013):1–7. doi:10.1038/ncomms3362. http://dx.doi.org/10.1038/ncomms3362.

Krebs R, Linnenbrink M, Guenther A. 2019. Validating standardised personality tests under semi-natural conditions in wild house mice (Mus musculus domesticus). Ethology. 125(11):761–773. doi:10.1111/eth.12930.

Kruuk LEB. 2004. Estimating genetic parameters in natural populations using the “animal model.” Philos Trans R Soc B Biol Sci. 359(1446):873–890. doi:10.1098/rstb.2003.1437.

Lynch M, Walsh B. 1998. Genetics and analysis of quantitative traits. Sinauer, Sunderland, MA.

Martin JGA, Nussey DH, Wilson AJ, Réale D. 2011. Measuring individual differences in reaction norms in field and experimental studies: A power analysis of random regression models. Methods Ecol Evol. 2(4):362–374. doi:10.1111/j.2041-210X.2010.00084.x.

Martin JGA, Réale D. 2008. Temperament, risk assessment and habituation to novelty in eastern chipmunks, Tamias striatus. Anim Behav. 75(1):309–318. doi:10.1016/j.anbehav.2007.05.026.

McCowan LSC, Mainwaring MC, Prior NH, Griffith SC. 2015. Personality in the wild zebra finch: Exploration, sociality, and reproduction. Behav Ecol. 26(3):735–746. doi:10.1093/beheco/aru239.

Mettke-Hofmann C, Lorentzen S, Schlicht E, Schneider J, Werner F. 2009. Spatial neophilia and spatial neophobia in resident and migratory warblers (Sylvia). Ethology. 115(5):482–492. doi:10.1111/j.1439-0310.2009.01632.x.

Miller JJ. 1977. Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. Ann Stat. 5(4):746–762.

Mitchell DJ, Houslay TM. 2021. Context-dependent trait covariances: how plasticity shapes behavioral syndromes. Behav Ecol. 32(1):25–29. doi:10.1093/beheco/araa115.

Moiron M, Araya-Ajoy YG, Mathot KJ, Mouchet A, Dingemanse NJ. 2019. Functional relations between body mass and risk-taking behavior in wild great tits. Behav Ecol. 30(3):617–623. doi:10.1093/beheco/ary199. https://academic.oup.com/beheco/article/30/3/617/5292535.

Moiron M, Laskowski KL, Niemelä PT. 2020. Individual differences in behaviour explain variation in survival: a meta-analysis. Ecol Lett. 23(2):399–408. doi:10.1111/ele.13438.

Mouchet A, Dingemanse NJ. 2021. Data from: A quantitative genetics approach to validate lab versus field-based behavior in novel environments. Behav Ecol. https://doi.org/10.5061/dryad.ht76hdrd0

Mutzel A, Dingemanse NJ, Araya-Ajoy YG, Kempenaers B. 2013. Parental provisioning behaviour plays a key role in linking personality with reproductive success. Proc R Soc B Biol Sci. 280(1764):1–9. doi:10.1098/rspb.2013.1019.

Nakagawa S, Schielzeth H. 2010. Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. Biol Rev. 85(4):935–956. doi:10.1111/j.1469-185X.2010.00141.x.

Neale MC, Eaves LJ, Kendler KS, Hewitt JK. 1989. Bias in correlations from selected samples of relatives: The effects of soft selection. Behav Genet. 19(2):163–169. doi:10.1007/BF01065901.

Nicolaus M, Brommer JE, Ubels R, Tinbergen JM, Dingemanse NJ. 2013. Exploring patterns of variation in clutch size-density reaction norms in a wild passerine bird. J Evol Biol. 26(9):2031–2043. doi:10.1111/jeb.12210.

Nicolaus M, Mathot KJ, Araya-Ajoy YG, Mutzel A, Wijmenga JJ, Kempenaers B, Dingemanse NJ. 2015. Does coping style predict optimization? An experimental test in a wild passerine bird. Proc R Soc B Biol Sci. 282(1799). doi:10.1098/rspb.2014.2405.

Nicolaus M, Tinbergen JM, Bouwman KM, Michler SPM, Ubels R, Both C, Kempenaers B, Dingemanse NJ. 2012. Experimental evidence for adaptive personalities in a wild passerine bird. Proc R Soc B Biol Sci. 279(1749):4885–4892. doi:10.1098/rspb.2012.1936.

Niemelä PT, Dingemanse NJ. 2014. Artificial environments and the study of “adaptive” personalities. Trends Ecol Evol. 29(5):245–247. doi:10.1016/j.tree.2014.02.007.

Niemelä PT, Dingemanse NJ. 2017. Individual versus pseudo-repeatability in behaviour: Lessons from translocation experiments in a wild insect. J Anim Ecol. 86(5):1033–1043. doi:10.1111/1365-2656.12688.

Niemelä PT, Dingemanse NJ. 2018a. On the usage of single measurements in behavioural ecology research on individual differences. Anim Behav. 145:99–105. doi:10.1016/j.anbehav.2018.09.012.

Niemelä PT, Dingemanse NJ. 2018b. Meta-analysis reveals weak associations between intrinsic state and personality. Proc R Soc B Biol Sci. 285(1873). doi:10.1098/rspb.2017.2823.

van Noordwijk AJ, de Jong G. 1986. Acquisition and Allocation of Resources: Their Influence on Variation in Life History Tactics. Am Nat. 128(1):137–142. doi:10.1086/284547.

Osborn A, Briffa M. 2017. Does repeatable behaviour in the laboratory represent behaviour under natural conditions? A formal comparison in sea anemones. Anim Behav. 123:197–206. doi:10.1016/j.anbehav.2016.10.036. http://dx.doi.org/10.1016/j.anbehav.2016.10.036.

van Overveld T, Careau V, Adriaensen F, Matthysen E. 2014. Seasonal- and sex-specific correlations between dispersal and exploratory behaviour in the great tit. Oecologia. 174(1):109–120. doi:10.1007/s00442-013-2762-0.

van Overveld T, Matthysen E. 2013. Personality and Information Gathering in Free-Ranging Great Tits. PLoS One. 8(2). doi:10.1371/journal.pone.0054199.

van de Pol M. 2012. Quantifying individual variation in reaction norms: How study design affects the accuracy, precision and power of random regression models. Methods Ecol Evol. 3(2):268–280. doi:10.1111/j.2041-210X.2011.00160.x.

van de Pol M, Wright J. 2009. A simple method for distinguishing within- versus between-subject effects using mixed models. Anim Behav. 77(3):753–758. doi:10.1016/j.anbehav.2008.11.006.

Quinn JL, Patrick SC, Bouwhuis S, Wilkin TA, Sheldon BC. 2009. Heterogeneous selection on a heritable temperament trait in a variable environment. J Anim Ecol. 78(6):1203–1215. doi:10.1111/j.1365-2656.2009.01585.x.

Réale D, Dingemanse NJ, Kazem AJN, Wright J. 2010. Evolutionary and ecological approaches to the study of personality. Philos Trans R Soc B Biol Sci. 365(1560):3937–3946. doi:10.1098/rstb.2010.0222.

Réale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ. 2007. Integrating animal temperament within ecology and evolution. Biol Rev. 82(2):291–318. doi:10.1111/j.1469-185X.2007.00010.x.

Roff DA. 1997. Evolutionary quantitative genetics. New York, NY: Chapman & Hall.

Roff DA. 2001. The threshold model as a general purpose normalizing transformation. Heredity (Edinb). 86(4):404–411. doi:10.1046/j.1365-2540.2001.00844.x.

Royauté R, Berdal MA, Garrison CR, Dochtermann NA. 2018. Paceless life? A meta-analysis of the pace-of-life syndrome hypothesis. Behav Ecol Sociobiol. 72(64):1–10.

Rudin FS, Simmons LW, Tomkins JL. 2019. Social cues affect quantitative genetic variation and covariation in animal personality traits. Evolution (N Y). 73(3):540–553. doi:10.1111/evo.13661.

Rudin FS, Tomkins JL, Simmons LW. 2018. The effects of the social environment and physical disturbance on personality traits. Anim Behav. 138:109–121. doi:10.1016/j.anbehav.2018.02.013. https://doi.org/10.1016/j.anbehav.2018.02.013.

Santostefano F, Wilson AJ, Niemelä PT, Dingemanse NJ. 2017. Behavioural mediators of genetic life-history trade-offs: A test of the pace-of-life syndrome hypothesis in field crickets. Proc R Soc B Biol Sci. 284(1864):15–17. doi:10.1098/rspb.2017.1567.

Schlichting CD. 2008. Hidden reaction norms, cryptic genetic variation, and evolvability. Ann N Y Acad Sci. 1133(1942):187–203. doi:10.1196/annals.1438.010.

Searle SR. 1961. Phenotypic, genetic and environmental correlations. Biometrics. 17(3):474–480.

Serrano-Davies E, O'Shea W, Quinn JL. 2017. Individual foraging preferences are linked to innovativeness and personality in the great tit. Behav Ecol Sociobiol. 71(11). doi:10.1007/s00265-017-2389-y.

Siepielski AM, Dibattista JD, Carlson SM. 2009. It's about time: The temporal dynamics of phenotypic selection in the wild. Ecol Lett. 12(11):1261–1276. doi:10.1111/j.1461-0248.2009.01381.x.

Siepielski AM, Gotanda KM, Morrissey MB, Diamond SE, Dibattista JD, Carlson SM. 2013. The spatial patterns of directional phenotypic selection. Ecol Lett. 16(11):1382–1392. doi:10.1111/ele.12174.

Sih A, Mathot KJ, Moirón M, Montiglio PO, Wolf M, Dingemanse NJ. 2015. Animal personality and state-behaviour feedbacks: A review and guide for empiricists. Trends Ecol Evol. 30(1):50–60. doi:10.1016/j.tree.2014.11.004.

Smit JAH, van Oers K. 2019. Personality types vary in their personal and social information use. Anim Behav. 151:185–193. doi:10.1016/j.anbehav.2019.02.002. https://doi.org/10.1016/j.anbehav.2019.02.002.

Smith BR, Blumstein DT. 2008. Fitness consequences of personality: A meta-analysis. Behav Ecol. 19(2):448–455. doi:10.1093/beheco/arm144.

Stamps JA. 2007. Growth-mortality tradeoffs and "personality traits" in animals. Ecol Lett. 10(5):355–363. doi:10.1111/j.1461-0248.2007.01034.x.

Stram D, Lee JW. 1994. Variance Components Testing in the Longitudinal Mixed Effects Model. Biometrics. 50(4):1171–1177.

Stuber EF, Araya-Ajoy YG, Mathot KJ, Mutzel A, Nicolaus M, Wijmenga JJ, Mueller JC, Dingemanse NJ. 2013. Slow explorers take less risk: A problem of sampling bias in ecological studies. Behav Ecol. 24(5):1092–1098. doi:10.1093/beheco/art035.

Verbeek MEM, Drent PJ, Wiepkema PR. 1994. Consistent individual differences in early exploratory behaviour of male great tits. Anim Behav. 48(5):1113–1121. doi:10.1006/anbe.1994.1344.

Westneat DF, Hatch MI, Wetzel DP, Ensminger AL. 2011. Individual variation in parental care reaction norms: Integration of Personality and plasticity. Am Nat. 178(5):652–667. doi:10.1086/662173.

Wilson AJ, Réale D, Clements MN, Morrissey MM, Postma E, Walling CA, Kruuk LEB, Nussey DH. 2010. An ecologist's guide to the animal model. J Anim Ecol. 79(1):13–26. doi:10.1111/j.1365-2656.2009.01639.x.

Wolf M, Van Doorn GS, Leimar O, Weissing FJ. 2007. Life-history trade-offs favour the evolution of animal personalities. Nature. 447(7144):581–584. doi:10.1038/nature05835.

Wolf M, Weissing FJ. 2010. An explanatory framework for adaptive personality differences. Philos Trans R Soc B Biol Sci. 365(1560):3959–3968. doi:10.1098/rstb.2010.0215.

Wolf M, Weissing FJ. 2012. Animal personalities: Consequences for ecology and evolution. Trends Ecol Evol. 27(8):452–461. doi:10.1016/j.tree.2012.05.001. http://dx.doi.org/10.1016/j.tree.2012.05.001.

Yuen CH, Pillay N, Heinrichs M, Schoepf I, Schradin C. 2016. Personality traits are consistent when measured in the field and in the laboratory in African striped mice (Rhabdomys pumilio). Behav Ecol Sociobiol. 70(8):1235–1246. doi:10.1007/s00265-016-2131-1. http://dx.doi.org/10.1007/s00265-016-2131-1.

SUPPORTING INFORMATION FOR CHAPTER 4

Appendix S1. Experiments

To control for potential effects of various experiments conducted in the population, we included a 4-level fixed factor brood size manipulation (enlarged, reduced, control or unmanipulated in years 2010 and 2011 (Nicolaus et al. 2012) vs. unmanipulated in all other years) into all models described in the Main Text. Sample sizes of, respectively, enlarged, reduced, control and unmanipulated levels were for laboratory-based behavior n=23, 20, 8 and 4253 broods and for field-based behavior n=115, 109, 102 and 3978 broods. We also included a 3-level fixed effect factor perceived predation risk (broadcast of predator vs. non-predator sounds during the breeding seasons 2012 and 2013 and during a month in winter 2014 (Abbey-Lee, Kaiser, et al. 2016; Abbey-Lee, Mathot, et al. 2016) vs. unmanipulated in all other years). Sample sizes of, respectively, treatment, control and unmanipulated levels were for laboratory-based behavior n=206, 203 and 3895 broods and for field-based behavior n=339, 369 and 3596 broods. Finally, we included a 3-level fixed effect factor supplementary feeding (supplementary feeding outside the breeding season vs. control in years 2015 to 2019 (detailed below), vs. unmanipulated in all other years). Sample sizes of, respectively, treatment, control and unmanipulated levels were for laboratory-based behavior n=246, 300 and 3758 broods and for field-based behavior n=773, 745 and 2786 broods. For each of these factors, we set unmanipulated as the reference category.

We conducted the supplementary feeding experiment during four consecutive seasons, taking place from mid-July of a calendar year until the end of March of the following calendar year. Supplementary food was thus not provided during the breeding phase (April-early July). In each of the treated plots, four feeders were placed such that they covered the entire plot. The experiment was conducted from July 2015 till March 2019, following a partial cross-over design. Within a year, supplementary food (sunflower seeds and fat balls) was provided in half of the plots while the other half was kept as control (untreated, not supplemented). The following year, half of the treated (vs. control) plots were switched to control (vs. treated), and the other half was again treated (vs. kept as control). All plots thus received both the treatment and control treatment twice over the same four-year period.

Appendix S2. Sampling bias

We tested whether the population sampled in winter was biased towards specific behavioral types as an explanation for failure to find correlations between laboratory- and field-based behavior. We did so by focusing on birds whose field-based behavior was assessed in a given breeding season and determining for each individual whether it was captured in the previous winter as part of our roosting inspection. We then tested whether field-based behavior differed between birds that were (1) versus were not (0) captured in the previous winter. Any such effects result from the combined influences of within- and among-individual effects of capture (van de Pol and Wright 2009). The former reflects within-individual changes (plasticity) in field-based behavior resulting from capture in the previous winter. The latter instead results from selective appearance of individuals with respect to their intercept for field-based behavior (i.e. "average" behavior or "behavioral type"), particularly if the effect differs significantly from the within-individual effect (detailed in (Sprau et al. 2017)).

We thus expanded our univariate model (Table S1 and S3) for field-based behavior by adding winter capture as an additional fixed effect. For each field-based activity score of an individual, winter capture was attributed a "1" if it had been caught in the winter preceding the focal field-based assay and a "0" if not. To acknowledge that effects of winter capture on field-based behavior might be attributable to either among-individual (i.e., selective sampling) or within-individual effects (i.e., phenotypic plasticity), we disentangled the two by applying within-subject centering approaches (van de Pol and Wright 2009). To estimate the among-individual effect, we fitted the mean value for winter capture (i.e., the proportion of winters the focal bird was captured) for each individual over all its observations. To estimate the within-individual effect, we subtracted this mean from each observation (0 versus 1). Both were fitted as predictors into the model.

Records of birds captured as breeders that were caught (versus not caught) during roosting in the preceding winter were associated with field-based activity scores that were, on average, 15% lower. This lower behavioral score was due to both among- and within-individual effects (Fig S1, Table S4). Lack of differences in within- versus among-individual effects indicate that the latter effect may also simply result from within-individual plasticity. This is not the case when among- versus within-individual effects differ statistically (Sprau et al. 2017). In such cases two distinct mechanisms are at play rather than just one. We therefore reformulated the model to test the difference (Δ) between among- and within-individual effects (van de Pol and Wright 2009). A difference between the among- versus within-individual effect was supported

(Table S4). Thus, over and above a plastic response (non-zero within-individual effect), there was evidence for sampling bias. Among-individual effects were stronger than their within-individual counterparts, implying that birds with higher intercept values for field-based behavior (i.e. more active birds) were less likely to be captured, as illustrated in Fig. S1.

Fig. S1. A reaction norm visualization illustrating the joint occurrence of within- and among-individual effects of winter capture on field-based activity behavior in a novel environment. The dashed-line represents the relationship between the average behavior of an individual and probability of winter capture frequency (i.e., the among-individual effect resulting from biased sampling). Each blue dot represents the average behavior of an individual and each blue line its behavioral reaction norm (i.e., its within-individual response to capture resulting from reversible plasticity). The plotted slopes match model estimates presented in Table S4.

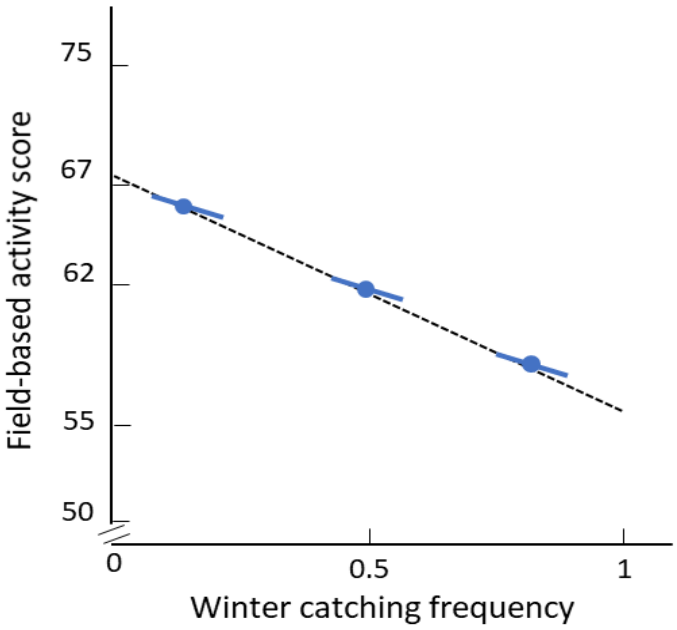


Table S1. Sources of variation of laboratory-based activity behavior in a novel environment. Fixed effects estimates (β) of univariate mixed-effect animal models are presented with standard errors (se). Conditional F-statistics (F-cond) with the number of degrees of freedom (df) and significance of the effects (p-value) are also presented.

Fixed Effects	β (se)	F-cond$_{(df)}$	p-value
(Intercept)	12.29 (1.60)	200.70$_{(1)}$	<0.001
Sex*	1.19 (0.62)	3.75$_{(1)}$	0.05
Sequence		13.18$_{(6)}$	<0.001
Sequence 1	8.47 (1.34)		
Sequence 2	11.27 (1.50)		
Sequence 3	16.05 (2.06)		
Sequence 4	7.40 (3.91)		
Sequence 5	15.12 (5.43)		
Sequence 6	11.84 (5.47)		
Test interval†	-0.01 (0.003)	6.27$_{(1)}$	0.01
Test time‡	0.10 (0.34)	0.09 $_{(1)}$	0.77
Brood size Manipulation§		3.07$_{(3)}$	0.03
+0 nestlings	2.42 (4.19)		
-3 nestlings	1.41 (2.52)		
+3 nestlings	7.62 (2.53)		
Feeding Manipulation§		3.07$_{(2)}$	0.08
Food control	3.80 (2.37)		
Food treated	5.36 (2.33)		
Perceived risk manipulation§		2.42$_{(2)}$	0.12
Predation control	5.71 (2.62)		
Predation treated	4.44 (2.62)		

*Mean-sex centered (males coded as -0.5, females as 0.5); the intercept is thus for the average sex.

†Difference between the first (sequence 0) and subsequent tests.

‡Time since sunrise and mean-centered.

§Difference between non-experimental years and the different levels of the experiment

Table S2. Sources of variation of field-based activity behavior in a novel environment. Fixed effects estimates (β) of univariate mixed-effects animal models are presented with their standard errors (se). Conditional F-statistics (F-cond) with the number of degrees of freedom (df) and significance of the effects (p-value) are also presented.

Fixed Effects	β (se)	F-cond$_{(df)}$	p-value
(Intercept)	65.88 (2.01)	2270.00$_{(1)}$	<0.001
Sex*	8.04 (0.77)	108.20$_{(1)}$	<0.001
Sequence		11.54$_{(7)}$	<0.001
Sequence 1	-1.1 (1.93)		
Sequence 2	-5.28 (2.03)		
Sequence 3	-12.73 (2.38)		
Sequence 4	-14.66 (3.18)		
Sequence 5	-21.03 (4.97)		
Sequence 6	-22.43 (17.24)		
Sequence 7	-24.93 (17.16)		
Test interval†	-0.01 (0)	1.77$_{(1)}$	0.14
Test time‡	0.50 (0.17)	8.71$_{(1)}$	0.003
Brood Size Manipulation§		0.41$_{(3)}$	0.74
+0 nestlings	1.02 (2.34)		
-3 nestlings	1.48 (2.25)		
+3 nestlings	-1.15 (2.23)		
Feeding manipulation§		0.46$_{(2)}$	0.64
Food control	2.21 (2.6)		
Food treated	2.5 (2.6)		
Perceived risk manipulations§		0.24$_{(2)}$	0.79
Predation control	1.03 (3.16)		
Predation treated	1.82 (3.16)		

*Mean-sex centered (males coded as -0.5, females as 0.5); the intercept is thus for the average sex.

†Difference between the first (sequence 0) and subsequent tests.

‡Time since sunrise and mean-centered.

§Difference between non-experimental years and the different levels of the experiment.

Table S3. Sources of variation of laboratory- and field-based behavior in a model checking whether sequence effect estimates (Table S1 and S2) were biased by categories with few data points. We achieved this by grouping sequences 3-6 for the laboratory, and sequences 4-7 for the field assay prior to re-analysis, which yielded similar estimates. The fixed and random effects structures were identical to the models presented in Tables S1, S2 and S3. Fixed effects estimates (β) of univariate mixed-effects animal models are presented with their standard errors (se). Conditional F-statistics (F-cond) with the number of degrees of freedom (df) and significance of the effects (p-value) are also presented.

	Laboratory-based behavior			**Field-based behavior**		
Fixed Effects	β (se)	F-cond$_{(df)}$	p-value	β (se)	F-cond$_{(df)}$	p-value
(Intercept)	13.02 (1.63)	199.50$_{(1)}$	<0.001	65.89 (2.01)	2274.00$_{(1)}$	<0.001
Sex*	-1.28 (0.59)	4.66$_{(1)}$	0.03	8.05 (0.77)	108.60$_{(1)}$	<0.001
Sequence†		26.07$_{(3)}$	<0.001		19.66$_{(4)}$	<0.001
Sequence 1	7.87 (1.31)			-1.03 (1.94)		
Sequence 2	11.19 (1.5)			-5.21 (2.03)		
Sequence 3	15.54 (1.97)			-12.67 (2.39)		
Sequence 4	-			-16.13 (2.95)		
Test interval	-0.01 (0)	5.28$_{(1)}$	0.02	-0.01 (0)	1.89$_{(1)}$	0.17
Test time ‡	-0.07 (0.33)	0.05 $_{(1)}$	0.83	0.51 (0.17)	8.99$_{(1)}$	0.003
Brood Size Manipulation§		2.47$_{(3)}$	0.06			
+0 nestlings	0.38 (3.54)			1.04 (2.34)	0.42$_{(3)}$	0.74
-3 nestlings	1.24 (2.29)			1.48 (2.25)		
+3 nestlings	5.97 (2.2)			-1.16 (2.23)		
Feeding manipulation§		2.78$_{(2)}$	0.09			
Food control	3.9 (2.41)			2.19 (2.6)	0.44$_{(2)}$	0.65
Food treated	5.31 (2.38)			2.47 (2.6)		
Perceived risk manipulation§		2.55$_{(2)}$	0.11			
Predation control	5.8 (2.67)			1.08 (3.16)	0.24$_{(2)}$	0.79
Predation treated	4.25 (2.67)			1.85 (3.17)		

*Mean-sex centered (males coded as -0.5, females as 0.5); the intercept is thus for the average sex.

†Difference between the first (sequence 0) and subsequent tests.

‡Time since sunrise and mean-centered.

§Difference between non-experimental years and the different levels of the experiment.

Table S4. Variance components of laboratory and field-based activity behavior in novel environments of univariate mixed-effects animal models. Variance estimates (σ^2) are presented with their standard errors (se). Statistical significance (p-value) was tested using log-likelihood ratio tests, from which χ^2-values were derived and considering a mixture of chi-square distributions with 0 and 1 degree of freedom ($\chi^2 0/1$). Repeatability (R) and heritability (h^2, genetic variance in column R) are also presented with their standard errors (se).

	Laboratory-based behavior				Field-based behavior			
Random effects	σ^2 (se)	$\chi^2_{0/1}$	p-value	R (se)	σ^2 (se)	$\chi^2_{0/1}$	p-value	R (se)
Year	7.63 (4.95)	30.04	<0.001	0.06 (0.04)	10.82 (6.61)	29.94	<0.001	0.02 (0.01)
Plot	2.59 (1.86)	5.29	0.01	0.02 (0.01)	7.83 (4.2)	21.62	<0.001	0.02 (0.01)
Plot-Year	2.54 (3.21)	0.78	0.19	0.00 (0.03)	0.00 (0.00)	0.00	0.50	0.00 (0.004)
Date	0.00 (0.00)	0.00	0.50	0.02 (0.03)	9.26 (3.31)	13.95	<0.001	0.02 (0.007)
Individual variance (V_I)	44.29 (5.28)	87.19	<0.001	0.35 (0.04)	164.05 (11.72)	300.86	<0.001	0.37 (0.02)
Genetic variance (V_A)	6.75 (12.25)	0.39	0.26	0.06 (0.10)	101.6 (24.49)	19.47	<0.001	0.23 (0.05)
Permanent environmental (V_{PE})	35.57 (12.90)	9.52	0.001	0.30 (0.11)	62.66 (24.37)	8.53	0.002	0.14 (0.05)
Residual	63.62 (4.50)	-	-	-	252.83 (9.64)	-	-	-
Total phenotypic variance (V_P)	117.70 (4.51)	-	-	-	445.18 (10.44)	-	-	-

Table S5. Analysis of sampling bias in roosting captures with respect to field-based behavior. Expanded univariate mixed-effects animal models fitting among- and within-individual effects of birds caught in the winter preceding a focal field-based test. We also present the difference between the within- and among-individual effects based on a reformulated model (van de Pol and Wright 2009). Fixed effects estimates (β) of models are presented with their standard errors (se). Conditional F-statistics (F-cond) with the number of degrees of freedom (df) and significance of the effects (p-value) are also presented.

Fixed Effects	β (se)	F-cond	p-value
(Intercept)	67.29 (1.97)	$2449_{(1)}$	<0.001
Sex*	7.77 (0.75)	$105.90_{(1)}$	<0.001
Sequence†		$8.41_{(7)}$	<0.001
Sequence 1	1.65 (1.92)		
Sequence 2	-1.63 (2.02)		
Sequence 3	-8.24 (2.38)		
Sequence 4	-10.41 (3.15)		
Sequence 5	-16.52 (4.91)		
Sequence 6	-17.43 (16.97)		
Sequence 7	-20.59 (16.9)		
Test interval	0 (0.00)	$1.26_{(1)}$	0.26
Test time‡	0.54 (0.17)	$10.75_{(1)}$	0.001
Brood Size Manipulation§		$0.56_{(3)}$	0.64
+0 nestlings	1.60 (2.3)		
-3 nestlings	1.85 (2.21)		
+3 nestlings	-0.89 (2.19)		
Feeding manipulation§		$0.53_{(2)}$	0.60
Food control	2.24 (2.53)		
Food treated	2.60 (2.52)		
Perceived risk manipulation§		$0.42_{(2)}$	0.66
Predation control	1.15 (3.07)		
Predation treated	2.24 (3.08)		
Roosted previous winter			
Among-individual effect	-11.75 (1.11)	$111.30_{(1)}$	<0.001
Within-individual effect	-8.42 (1.15)	$53.58_{(1)}$	<0.001
Δ (Among-Within)	1.36 (0.62)	$4.86_{(1)}$	0.03

*Mean-sex centered (males coded as -0.5, females as 0.5); the intercept is thus for the average sex.

†Difference between the first (sequence 0) and subsequent tests.

‡Time since sunrise and mean-centered.

§Difference between non-experimental years and the different levels of the experiment.

Table S6. Estimates of a correlation between two traits in simulated data with versus without sampling bias in one of the traits. The correlation between the two traits in the biased sample was compared to the (simulated) true correlation (ranging from 0.1 to 0.5). The detection of the realized correlation in the data was tested using Pearson's correlation test. Parameters of the simulated samples were based on our empirical data with one measure per individual for simplicity. We first generated a population of 2047 individuals with two traits, corresponding to the field- and laboratory-based behavior, of known correlation using a multivariate normal distribution. Trait means and variances were extracted from the univariate model (μ_{field} (σ^2) = 65.88 (164.05), μ_{lab} (σ^2) = 12.29 (44.29)). We then generated a biased sample of 515 individuals, such that individuals with a field-based score above the mean of the biased sample (μ_{biased}=55.54, Table S4) had 9.6% chance of being sampled, which corresponded to the proportion of breeders with field-based scores above 55.54 not captured in winter. We then recalculated the correlation between the two traits in this biased population sample. The table demonstrates that the magnitude of the sampling bias in our data cannot readily explain the lack of correlation between field- and laboratory-based behavior, as the bias of this magnitude had very little effect on the estimated correlation.

True correlation	0.5	0.4	0.3	0.2	0.1
p-value	<0.001	<0.001	<0.001	<0.001	<0.001
95%CI	0.47;0.53	0.37;0.43	0.26;0.33	0.16;0.24	0.06; 0.14
Estimated correlation*	0.49	0.42	0.33	0.17	0.14
p-value	<0.001	<0.001	<0.001	<0.001	0.001
95%CI	0.42;0.55	0.34;0.48	0.25; 0.40	0.08; 0.26	0.06;0.22

*In the presence of sampling bias

SI REFERENCES

Abbey-Lee RN, Kaiser A, Mouchet A, Dingemanse NJ. 2016. Immediate and carry-over effects of perceived predation risk on communication behavior in wild birds. Behav Ecol. 27(3). doi:10.1093/beheco/arv210.

Abbey-Lee RN, Mathot KJ, Dingemanse NJ. 2016. Behavioral and morphological responses to perceived predation risk: A field experiment in passerines. Behav Ecol. 27(3):857–864. doi:10.1093/beheco/arv228.

Nicolaus M, Tinbergen JM, Bouwman KM, Michler SPM, Ubels R, Both C, Kempenaers B, Dingemanse NJ. 2012. Experimental evidence for adaptive personalities in a wild passerine bird. Proc R Soc B Biol Sci. 279(1749):4885–4892. doi:10.1098/rspb.2012.1936.

van de Pol M, Wright J. 2009. A simple method for distinguishing within- versus between-subject effects using mixed models. Anim Behav. 77(3):753–758. doi:10.1016/j.anbehav.2008.11.006.

Sprau P, Mouchet A, Dingemanse NJ. 2017. Multidimensional environmental predictors of variation in avian forest and city life histories. Behav Ecol. 28(1):59–68. doi:10.1093/beheco/arw130. https://academic.oup.com/beheco/article-lookup/doi/10.1093/beheco/arw130.

General Discussion

Behavioural variation has long fascinated behavioural ecologists, and certainly since the discovery of repeatable among-individual differences in behaviour within populations, called animal personality (Wilson 1998). Behaviour is both genetically and environmentally determined. Thus, behaviour has the potential to evolve by means of natural selection and be adaptive. Natural selection is expected to remove certain type of traits that are detrimental to an individual's fitness and to maintain traits that are advantageous (Darwin 1859). As such, selection pressures are expected to favour the phenotype best adapted to the environmental conditions and to ultimately erode variation. Moreover, behaviour is highly plastic such that individuals can adjust their behaviour quickly to changes in their environment (Lynch and Walsh 1998). How personalities coexist despite the action of natural selection and the highly labile nature of behaviour is intriguing. The maintenance of behavioural variation may increase abilities of populations to adapt to variable environments. Identifying the mechanisms involved in the maintenance of individual variation in behaviour will help better understanding how populations persist and evolve in constantly changing environments.

Among-individual variation in behaviour has been proposed to be adaptive if the costs and benefits of different behavioural phenotypes vary with the (a)biotic environment (Frank and Slatkin 1990; Dingemanse and Réale 2005; Dingemanse and Wolf 2010). Three main mechanisms may be at play: frequency-dependent selection, heterogeneous selection and life-history trade-offs. Though theoretical models have been developed, few studies have empirically tested the role of these mechanisms in the coexistence of animal personalities and determined their ecological drivers. This PhD thesis contributed to this gap by focusing on life-history trade-offs and heterogeneous selection and testing ecological drivers of heterogeneous selection in great tit populations. I found that both mechanisms played a key role in maintaining alternative behavioural tactics within populations (Chapter 1 and 2) but failed in determining the ecological drivers of heterogeneous selection (Chapter 2 and 3; Box 1).

Differential resolution of life-history trade-offs

A major life-history trade-off that animals reproducing multiple times face is a trade-off between current and future reproduction. The resolution of this trade-off is generally regarded as investing into current reproduction at the expense of survival. Risk-prone individuals may have lower survival probabilities and, thus, lower future reproductive expectations. Consequently, risk-prone individuals are expected to invest more into current reproduction and risk-adverse individuals into future reproduction. However, empirical support for such differential resolution of trade-off between reproduction and survival among behavioural

Box 1. Brief overview of chapters' main results

Chapter 1: Personality-related age-dependent reproduction with fast explorers increasing and subsequently decreasing reproductive investment with age and slow explorers laying similar clutch sizes throughout their lives.

Chapter 2: Heterogeneous selection on exploration behaviour within and among great tit populations with stronger spatial than temporal variation in selection. Existence of phenotypic variation in all populations suggests that temporal variation, however, plays a key role in counteracting local adaptation favoured by spatial variation. Heterogeneous selection was not driven by density-dependence.

Chapter 3: Selection on activity in a novel environment did not vary with experimental manipulation of food availability. Increased food availability, however, amplified behavioural variation among-individuals, either by increasing levels of plasticity or survival of more active individuals.

Chapter 4: Laboratory- and field-based behaviour in novel environments do not correlate and differ in heritability. Different assays scoring seemingly similar behaviours may present different contexts that may elicit the expression of different behaviours. The population sampled for the laboratory-based test was biased toward more active individuals, and the winter capture had carry-over effects on the field-based behaviour scored in subsequent spring.

phenotypes is mixed (Mathot and Frankenhuis 2018; Royauté et al. 2018). Investing into current reproduction may not impact future survival but reproductive senescence instead. Indeed, in Chapter 1, we showed that fast explorers increased their clutch size in their first years and then decreased it later in age, while slower explorers laid similar clutches throughout their lifetime (Dingemanse et al. 2020). This finding suggests reproductive senescence in fast explorers. Interestingly, yearling birds produced similar clutch sizes regardless of their exploration behaviour. This study demonstrated that failing to account for age differences between individuals may lead to the conclusion that behavioural phenotypes do not differ in reproductive investment. Similarly, considering lifetime reproductive success instead of reproductive success per breeding attempt may also lead to such conclusions. Overall, it is important to consider that life-history trade-offs can be resolved in various ways and that intrinsic and extrinsic factors may influence the resolution of these trade-offs.

Consequences of heterogeneous selection on animal personality in metapopulations

Variation in ecological conditions are expected to cause selection to vary spatially across locations and temporally, for example, across years (Levene 1953; Gillespie 1974; Felsenstein 1976; Lande 1976; Siepielski et al. 2011; Siepielski et al. 2013). Spatial variation in selection should favour phenotypes locally adapted and thus population divergence (Grant and Price 1981; Foster and Endler 1999). Temporal variation in selection (or fluctuating selection)

instead should alternately favour different phenotypes within the same location (Haldane and Jayakar 1963; Hedrick 1976; Byers 2005). Thus, fluctuating selection may counteract population divergence. Though spatial and temporal variation co-occur and may have counteracting effects, with important implications for population adaptation, both processes have rarely been studied simultaneously. In a collaborative project, I assessed the relative importance of spatial and temporal variation in selection on exploration behaviour, using fitness and behavioural data from multiple West-European great tit populations (Chapter 2). I found that selection varied primarily spatially among populations, but also within populations among forest plots. Simultaneously, selection varied temporally. Though temporal variation in selection was of a lesser extent, this process must play a key role in counteracting population divergence promoted by spatial variation and maintaining behavioural variation in all populations. Ecological factors synchronised at the continent scale, such as beech masting, may drive large-scale temporal variation. However, temporal variation was population-specific, suggesting that ecological conditions characteristic to each population also played a role. Contrary to previous findings, density-dependence did not drive heterogeneous selection on exploration behaviour.

This study overall suggests that ecological factors acting both at macro- and micro-spatial scale shape behaviour and the ability of populations to respond to global and local spatiotemporal environmental changes. Genetic studies have demonstrated that European great tit populations are all genetically interconnected (Kvist et al. 2003). Our study, by estimating selection at the metapopulation level, implies that personalities mediate this metapopulation structure and may influence population expansion and adaptation in interaction with ecological factors (Duckworth and Badyaev 2007; Duckworth 2008; Edelaar and Bolnick 2012; Mouchet et al. 2021). We suggest that individuals disperse and settle non-randomly with respect to their behavioural phenotype (Clobert et al. 2009; Cote et al. 2010; Edelaar and Bolnick 2012). Non-random dispersal may be a passive process induced by ecological factors (e.g., winter temperature) or phenotypic traits (e.g., morphology, physiology) constraining dispersal differently among phenotypes (Bowler and Benton 2005; Lemoine et al. 2016). Conversely, non-random dispersal may be an active process, whereby individuals select environments that best suit their phenotypes, either based on abiotic (e.g., food resources, habitat cover) or biotic (e.g., predation risk, phenotypes of conspecifics) factors. However, exploration behaviour among first-year birds varied little among populations. Matching habitat choice thus, might not be a primary process occurring in this species, for example, because of high costs of dispersal or inaccurate information about the settling environment. Alternatively, temporal variation in

selection is strong enough to counteract the directional matching habitat choice promoted by spatial variation in selection (Edelaar et al. 2017). Selection on other phenotypic traits (other behaviours or morphological, physiological traits) may also be opposite to selection on exploration behaviour and favour phenotypic variation within populations (Cote et al. 2013). The next challenges are to determine i) whether settlement of young great tits differing in exploration behaviour is strongly habitat-specific, if so, ii) whether this phenotype-environment matching fluctuates across years and iii) whether multivariate selection on behaviour shows as strong spatial variation.

Ecological drivers of heterogeneous selection

Various ecological factors vary spatiotemporally and may affect individual fitness through changes in availability of and access to resources. Changes in resources may affect behavioural phenotypes differently as they differ, for example, in their competitive abilities, foraging strategies, risk of predation, social information use. Variation in abiotic and biotic factors is thus expected to drive heterogeneous selection on personalities. Population density and food availability are two factors varying spatiotemporally in most populations and affecting access to food resources essential for an individual's fitness. Both factors are thus expected to be important drivers of heterogeneous selection on personalities.

A previous study had investigated density-dependent viability selection on exploration behaviour in our population of great tits in Germany (Nicolaus et al. 2016). This study found that, indeed, population density affected behavioural phenotypes differently: fast explorers survived better in low breeding densities and slow explorers survived better in high densities. We were interested in knowing whether personality-related density-dependence was a mechanism generally occurring in all great tit populations. We thus tested whether our finding of heterogeneous selection on exploration behaviour in five great tit populations was driven by variation in breeding density (Chapter 2). Because these analyses were observational, we also designed an experiment to test the causal effect of density on heterogeneous selection (Chapter 3). Breeding density generally positively covaries with food availability: increases in food availability induce density increases through higher survival, recruitment rate and/or immigration rate (Perrins 1965; Balen 1980; Perdeck et al. 2000). Consequently, density-dependence of fitness may be a direct or an indirect effect through food availability. We thus manipulated food availability outside the breeding season in our great tit population with the expectation that variation in food availability would cause variation in breeding density and in fecundity selection on exploration behaviour (Chapter 3).

Both our studies yielded unexpected results. First, heterogeneous selection on exploration behaviour was not driven by breeding density. Second, fecundity selection on behaviour did not vary with our manipulation of food availability. In both cases, the factor of interest may be driving selection on behaviour, but its effect be counteracted by another factor. As expected, our food manipulation increased breeding densities (Chapter 3). Abundant food may favour fast explorers because they rely more on social information and superficially explore the environment (Groothuis and Carere 2005). In turn, the associated high density may disadvantage fast explorers that are also more aggressive, if costs of aggressiveness are too high in highly competitive contexts (Verbeek et al. 1999; Careau and Garland 2012; Niemelä and Dingemanse 2018; Mathot et al. 2019). It may explain that the selection gradient on exploration behaviour was overall close to zero. Future studies should disentangle the effects of food availability and population density on selection on behaviour to better understand the role of these factors on behavioural variation. Interestingly, high food availability context increased among-individual variation in behaviour and the average level of behaviour. Thus, some environmental factors amplify differences in behaviour among individuals. This may be due to increased levels of plasticity or increased survival, recruitment or immigration rate of faster explorers. Survival analyses and immigrant behaviour would enable us to distinguish both mechanisms. Estimating cross-context plasticity would prove much more challenging as a same individual should be assayed for behaviour in both contexts within the same season.

Validation of behavioural assays quantifying individual variation

In all chapters of this thesis, we used behaviour in novel environments as proxy for risk-taking behaviour. However, we did not assay behaviour using the same experimental design. In Chapter 2, we used behaviour scored in the laboratory because this assay was used in multiple great tit populations. This allowed us to infer differences in selection pressures on personality among populations. It also allowed us to draw general rather than population-specific conclusions on variation in selection in great tit populations (Mouchet et al. 2021). In Chapter 1 and 3, however, we used behaviour scored in the field because this assay allowed us to avoid known sampling biases of the catching method used for the laboratory assay and to increase sample sizes (Biro and Dingemanse 2009; Stuber et al. 2013).

We assumed both designs assayed the same behaviour because both laboratory- and field-based behaviour had been shown to each correlate with other field-based risk-taking behaviours, such as aggressiveness and boldness (e.g., Hollander et al. 2008; Quinn et al. 2009; Amy et al. 2010; Mutzel et al. 2013). Thus, we assumed both laboratory- and field-based

behaviour represented similar behaviours under the same selection pressures. However, in Chapter 4, we showed that birds do not express the same behaviour in both apparatuses (Mouchet and Dingemanse 2021). In the laboratory, birds may express exploration and fear, while in the field, they may express escape behaviour. Though multiple differences existed between the two assays, we suggested that the difference in behavioural expression may be due to the quasi absence versus presence of social cues. We also confirmed that the population sampled for the laboratory assay was a biased sample toward faster explorers. Finally, we demonstrated that the winter capture carried over to the behaviour assayed in spring. Chapter 4, thus, highlights the difficulty assaying behaviour in an unbiased and reproducible manner.

To investigate phenotypic variation and its underpinning mechanisms, it is essential to make sure that the population sample is unbiased with respect to phenotypes, either by using non-selective catching methods or combining multiple methods (Carter et al. 2013). We also need to bear in mind that animals may sort of memorise previous captures for extended periods of time (e.g., several months), which may affect their behaviour on subsequent captures. If all individuals are not subjected to the same capture events, it may be important to account for previous captures. The difficulty also lies in reproducing behavioural assays scoring seemingly similar behaviours (e.g., number of hops between perches). Animals may perceive any differences in assays as distinct contexts and respond by expressing different behaviours (Arvidsson et al. 2017; Rudin et al. 2019). These behaviours may be underpinned by different genes and be under different selection pressures. It is thus important to cross-validate behavioural assays before inferring they score the same behaviour than another seemingly similar assay and making biological predictions.

Importance of long-term and replication studies

Fully understanding individual variation in behaviour entails long-term and replication studies. The study of among-individual variation in behaviour requires large sample sizes to be able to characterise behavioural variation representative of the populations. It also requires scoring the same individual multiple times to determine its representative average behaviour and disentangle among- and within-individual variation (i.e., personality and plasticity) that may show different to opposite patterns (Mouchet and Dingemanse 2021; Sprau and Dingemanse 2017, Chapter 3). Forcefully understanding the adaptive causes and consequences of among-individual variation in behaviour additionally requires temporal and spatial replicates. Indeed, organisms live in variable environments that may select for different behavioural phenotypes.

Estimating selection pressures on behaviour in only one population will characterise evolutionary trajectories of behaviour specific to the environmental context of that population. A lack of spatial replicates may hide spatial variation in selection associated with other environmental contexts and thus draw an incomplete picture of how behaviour evolves, and animals adapt to their environment. Estimating selection in a short time window may hide fluctuating or rarer selection patterns associated with temporally changing environmental conditions and that may have important consequences for population dynamics and evolution (Grant and Grant 2002). Because it is challenging to monitor multiple populations in different locations, particularly at larger spatial scales, researchers may greatly benefit from collaborating and gathering the data from their individual populations. Such collaborative and long-term projects proved fruitful to achieve this PhD work on adaptive causes and ecological drivers of individual variation in behaviour in great tit populations. Collaborative projects, however, require using similar methods to score behaviour and standardising data. The growing current effort at gathering data from similar type of studies in unique, public and standardised databases will facilitate the fulfilment of more of such projects (e.g., Culina 2020).

Role of animal personalities in eco-evolutionary dynamics

Among-individual variation in behaviour implies that individuals use and interact with their environment nonrandomly. Individuals may specialise on certain food items, forage in specific time windows, disperse in different habitats, which in turn may affect intraspecific competition, mating opportunities, conspicuousness to predators or exposure to parasites (Dingemanse and De Goede 2004; Duckworth and Badyaev 2007; Svanbäck and Bolnick 2007; Gharnit et al. 2020; Rollins et al. 2021) Consequently, individuals may differ in their reproductive investment, as shown in Chapter 1, and in their response to environmental changes. In other words, behavioural phenotypes contribute differently to population dynamics, community structure and, overall, to ecosystem functioning. Among-individual variation in behaviour, or more generally, intraspecific phenotypic variation has important ecological consequences and should thus be considered in ecological and evolutionary studies. Understanding the consequences of behavioural variation may become especially important in a context of biodiversity loss, urbanisation and climate change.

Shifts in phenotypic composition of a population, for example through the action of heterogeneous selection (Chapter 3), should affect how and which resources are used, and thus alter the abiotic environment. Shifts in phenotypic composition should also alter interactions

between conspecifics and with competitor or predator heterospecifics, thus the dynamic and structure of populations and communities as the whole. This effect of evolutionary change (phenotypic composition through the action of natural selection) on ecology (population dynamics, community structure) may then feedback with the ecological change in turn affecting evolution (Fig.1). Phenotypic variation, including behavioural variation, should play a major role in such eco-evolutionary dynamics and thus in the ability of populations to adapt to different environments.

This PhD thesis contributed to understanding the role of adaptive mechanisms in among-individual variation in behaviour and their ecological drivers. This work therefore contributed to the understanding of eco-evolutionary dynamics between phenotypes and populations. This work forcefully considered temporal and spatial variation simultaneously and investigated the effects of two ecological drivers on selection. However, the effects of ecological drivers were studied separately though they may interact, or even counteract each other. Another important limit to this work is that it also used a single-trait and single-species approach. A comprehensive understanding of population adaptation to variable environments will require integrative studies that simultaneously consider several traits, multiple ecological drivers of selection and trophic interactions. Only these challenging approaches will embrace the complexity of natural environments and ecological interactions.

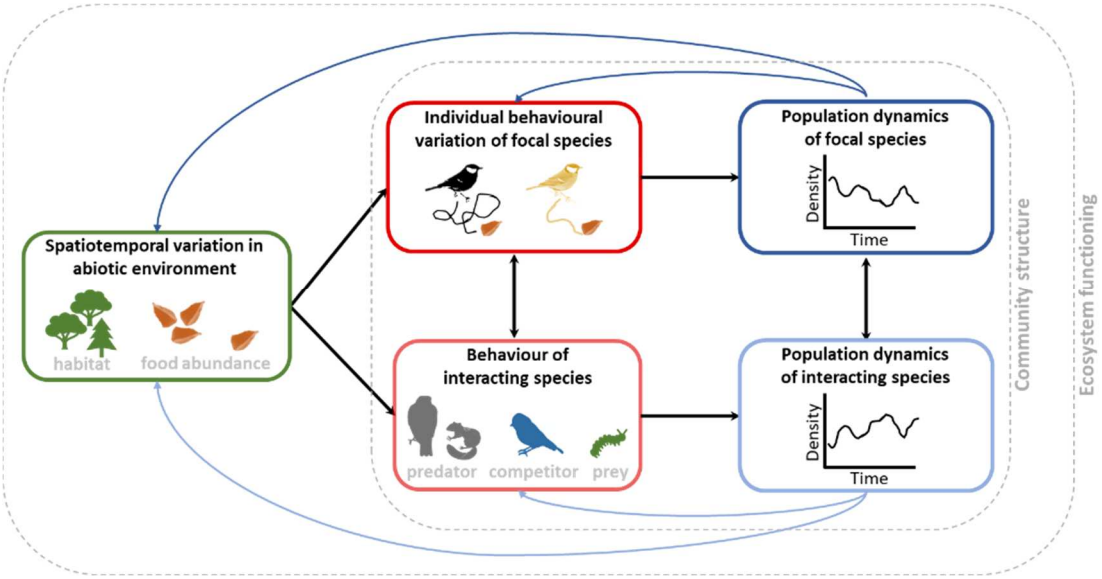


Fig. 1. Role of behaviour in eco-evolutionary dynamics. Variation in abiotic factors (green box) affects the behaviour of the focal species (dark red box), which in turn affects population dynamics (dark blue box). Changes in focal species' phenotypes affect phenotypes and population dynamics of interacting species (light red and blue boxes). Changes in population dynamics feedback on behaviour and ecology (blue arrows)

REFERENCES

Amy, M., Sprau, P., De Goede, P. & Naguib, M. (2010). Effects of personality on territory defence in communication networks: A playback experiment with radio-tagged great tits. Proc. R. Soc. B Biol. Sci., 277, 3685–3692.

Arvidsson, L.K., Adriaensen, F., van Dongen, S., De Stobbeleere, N. & Matthysen, E. (2017). Exploration behaviour in a different light: testing cross-context consistency of a common personality trait. Anim. Behav., 123, 151–158.

Balen, J.H. Van. (1980). Population Fluctuations of the Great Tit and Feeding Conditions in Winter. Ardea, 68, 143–164.

Biro, P.A. & Dingemanse, N.J. (2009). Sampling bias resulting from animal personality. Trends Ecol. Evol., 24, 63–66.

Bowler, D.E. & Benton, T.G. (2005). Causes and consequences of animal dispersal strategies: Relating individual behaviour to spatial dynamics. Biol. Rev. Camb. Philos. Soc., 80, 205–225.

Byers, D.L. (2005). Evolution in heterogeneous environments and the potential of maintenance of genetic variation in traits of adaptive significance. Genetica, 123, 107–124.

Careau, V. & Garland, T. (2012). Performance, personality, and energetics: Correlation, causation, and mechanism. Physiol. Biochem. Zool., 85, 543–571.

Carter, A.J., Feeney, W.E., Marshall, H.H., Cowlishaw, G. & Heinsohn, R. (2013). Animal personality: What are behavioural ecologists measuring? Biol. Rev., 88, 465–475.

Clobert, J., Le Galliard, J.F., Cote, J., Meylan, S. & Massot, M. (2009). Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. Ecol. Lett., 12, 197–209.

Cote, J., Clobert, J., Brodin, T., Fogarty, S. & Sih, A. (2010). Personality-dependent dispersal: Characterization, ontogeny and consequences for spatially structured populations. Philos. Trans. R. Soc. B Biol. Sci., 365, 4065–4076.

Cote, J., Fogarty, S., Tymen, B., Sih, A. & Brodin, T. (2013). Personality-dependent dispersal cancelled under predation risk. Proc. R. Soc. B Biol. Sci., 280.

Culina, A. (2020). Connecting the data landscape of long-term ecological studies: the SPI-Birds data hub. J. Anim. Ecol.

Darwin, C. (1859). On the origin of species. London.

Dingemanse, N.J. & De Goede, P. (2004). The relation between dominance and exploratory behavior is context-dependent in wild great tits. Behav. Ecol., 15, 1023–1030.

Dingemanse, N.J., Moiron, M., Araya-Ajoy, Y.G., Mouchet, A. & Abbey-Lee, R.N. (2020). Individual variation in age-dependent reproduction: Fast explorers live fast but senesce young? J. Anim. Ecol., 89, 601–613.

Dingemanse, N.J. & Réale, D. (2005). Natural selection and animal personality. Behaviour, 142, 1159–1184.

Dingemanse, N.J. & Wolf, M. (2010). Recent models for adaptive personality differences: A review. Philos. Trans. R. Soc. B Biol. Sci., 365, 3947–3958.

Duckworth, A. & Badyaev, A. V. (2007). Coupling of dispersal and aggression facilitates the rapid range expansion of a passrine bird. Proc. Natl. Acad. Sci., 104, 15017–15022.

Duckworth, R.A. (2008). Adaptive Dispersal Strategies and the Dynamics of a Range Expansion. Am. Nat., 172, S4–S17.

Edelaar, P. & Bolnick, D.I. (2012). Non-random gene flow: An underappreciated force in evolution and ecology. Trends Ecol. Evol., 27, 659–665.

Edelaar, P., Jovani, R. & Gomez-Mestre, I. (2017). Should I change or should I go? Phenotypic plasticity and matching habitat choice in the adaptation to environmental heterogeneity. Am. Nat., 190, 506–520.

Felsenstein, J. (1976). The theoritical population genetics of variable selection and migration. Annu. Rev. Genet., 10, 253–280.

Foster, S.A. & Endler, J.A. (1999). Geographic variation in behavior. Perspectives on evolutionary mechanisms. Oxford University Press, New York.

Frank, S.A. & Slatkin, M. (1990). Evolution in a Variable Environment. Am. Nat., 136, 244–260.

Gharnit, E., Bergeron, P., Garant, D. & Réale, D. (2020). Exploration profiles drive activity patterns and temporal niche specialization in a wild rodent. Behav. Ecol., 31, 772–783.

Gillespie, J. (1974). Polymorphism in patchy environments. Am. Nat., 108, 145–151.

Grant, P.R. & Grant, B.R. (2002). Unpredictable Evolution in a 30Year Study Darwin Finches, 296, 707–712.

Grant, P.R. & Price, T.D. (1981). Population variation in continuously varying traits as an ecological genetics problem. Am. Zool., 21, 795–811.

Groothuis, T.G.G. & Carere, C. (2005). Avian personalities: Characterization and epigenesis. Neurosci. Biobehav. Rev., 29, 137–150.

Haldane, J.B.S. & Jayakar, S.D. (1963). Polymorphism due to selection of varying direction. J. Genet., 58, 237–242.

Hedrick, P.W. (1976). Genetic variation in a heterogeneous environment. II. Temporal heterogeneity and directional selection. Genetics, 84, 145–157.

Hollander, F.A., Van Overveld, T., Tokka, I. & Matthysen, E. (2008). Personality and nest defence in the great tit (Parus major). Ethology, 114, 405–412.

Kvist, L., Martens, J., Higuchi, H., Nazarenko, A.A., Valchuk, O.P. & Orell, M. (2003). Evolution and genetic structure of the great tit (Parus major) complex. Proc. R. Soc. B Biol. Sci., 270, 1447–1454.

Lande, R. (1976). Natural Selection and Random Genetic Drift in Phenotypic Evolution. Evolution (N. Y)., 30, 314.

Lemoine, M., Lucek, K., Perrier, C., Saladin, V., Adriaensen, F., Barba, E., et al. (2016). Low but contrasting neutral genetic differentiation shaped by winter temperature in European great tits. Biol. J. Linn. Soc., 118, 668–685.

Levene, H. (1953). Genetic Equilibrium When More Than One Ecological Niche is Available. Am. Nat., 87, 331–333.

Lynch, M. & Walsh, B. (1998). Genetics and analysis of quantitative traits. Sinauer, Sunderland, MA.

Mathot, K.J., Dingemanse, N.J. & Nakagawa, S. (2019). The covariance between metabolic rate and behaviour varies across behaviours and thermal types: meta-analytic insights. Biol. Rev., 94, 1056–1074.

Mathot, K.J. & Frankenhuis, W.E. (2018). Models of pace-of-life syndromes (POLS): a systematic review. Behav. Ecol. Sociobiol., 72, 1–12.

Mouchet, A., Cole, E.F., Matthysen, E., Nicolaus, M., Quinn, J.L. & Roth, A.M. (2021). Heterogeneous selection on exploration behavior within and among West European populations of a passerine bird. Proc. Natl. Acad. Sci., 118, 1–6.

Mouchet, A. & Dingemanse, N.J. (2021). A quantitative genetics approach to validate lab-versus field-based behavior in novel environments. Behav. Ecol., 1–9.

Mutzel, A., Dingemanse, N.J., Araya-Ajoy, Y.G. & Kempenaers, B. (2013). Parental provisioning behaviour plays a key role in linking personality with reproductive success. Proc. R. Soc. B Biol. Sci., 280, 1–9.

Nicolaus, M., Tinbergen, J.M., Ubels, R., Both, C. & Dingemanse, N.J. (2016). Density fluctuations represent a key process maintaining personality variation in a wild passerine bird. Ecol. Lett.

Niemelä, P.T. & Dingemanse, N.J. (2018). Meta-analysis reveals weak associations between intrinsic state and personality. Proc. R. Soc. B Biol. Sci., 285, 1–7.

Perdeck, A.C., Visser, M.E. & van Balen, J.H. (2000). Great tit Parus major survival and the beech-crop cycle. Ardea, 88 (1), 99–108.

Perrins, C.M. (1965). Population fluctuations and clutch size in the great tit, Parus major L. J. Anim. Ecol., 34, 601–647.

Quinn, J.L., Patrick, S.C., Bouwhuis, S., Wilkin, T.A. & Sheldon, B.C. (2009). Heterogeneous selection on a heritable temperament trait in a variable environment. J. Anim. Ecol., 78, 1203–1215.

Rollins, R.E., Mouchet, A., Margos, G., Chitimia-Dobler, L., Fingerle, V., Becker, N.S., et al. (2021). Repeatable differences in exploratory behaviour predict tick infestation probability in wild great tits. Behav. Ecol. Sociobiol., 75.

Royauté, R., Anderson Berdal, M., Garrison, C.R. & Dochtermann, N.A. (2018). Paceless life? A meta-analysis of the pace-of-life syndrome hypothesis. Behav. Ecol. Sociobiol., 72, 1–10.

Rudin, F.S., Simmons, L.W. & Tomkins, J.L. (2019). Social cues affect quantitative genetic variation and covariation in animal personality traits. Evolution (N. Y)., 73, 540–553.

Siepielski, A.M., DiBattista, J.D., Evans, J.A. & Carlson, S.M. (2011). Differences in the temporal dynamics of phenotypic selection among fitness components in the wild. Proc. R. Soc. B Biol. Sci., 278, 1572–1580.

Siepielski, A.M., Gotanda, K.M., Morrissey, M.B., Diamond, S.E., Dibattista, J.D. & Carlson, S.M. (2013). The spatial patterns of directional phenotypic selection. Ecol. Lett., 16, 1382–1392.

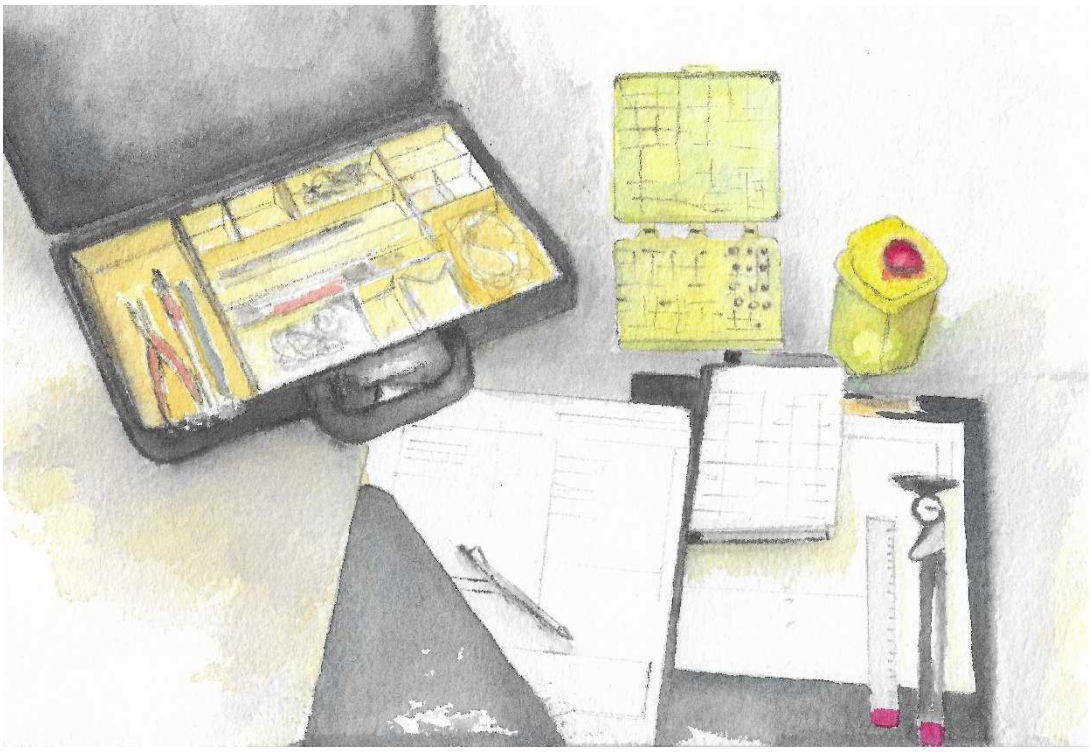
Stuber, E.F., Araya-Ajoy, Y.G., Mathot, K.J., Mutzel, A., Nicolaus, M., Wijmenga, J.J., et al. (2013). Slow explorers take less risk: A problem of sampling bias in ecological studies. Behav. Ecol., 24, 1092–1098.

Svanbäck, R. & Bolnick, D.I. (2007). Intraspecific competition drives increased resource use diversity within a natural population. Proc. R. Soc. B Biol. Sci., 274, 839–844.

Verbeek, M.E.M., De Goede, P., Drent, P.J. & Wiepkema, P.R. (1999). Individual behavioural characteristics and dominance in aviary groups of great tits. Behaviour, 136, 23–48.

Wilson, D.S. (1998). Adaptive single differences within single populations. Philos. Trans. R. Soc. London B, 353, 199–205.

Acknowledgements



I began this PhD thesis at a time when I started to accept that it was time to think about another career path. Many applications later and several years after obtaining my Masters, the news I was not expecting anymore broke on early 2017. I was just back from fieldwork that members of the group rushed into me to tell me that Niels (Dingemanse) was on the phone and wanted to talk to me. The grant application had been accepted! The news was then not fully unexpected as a few months earlier Niels had already asked me if I would still like to do a PhD. Of course, I would not miss such an opportunity! Four years later, this dissertation is achieved. It was a long and hard process, but I really enjoyed it and I am happy that I persevered. It is now time to thank everyone who contributed one way or another to this thesis.

I am grateful to my advisor Niels Dingemanse for trusting a "slow" personality could achieve this PhD thesis. You had already offered me the opportunity to work in your group as a technical assistant for some years, which allowed me to gain more experience in research, new skills and have responsibilities. This experience enabled me to gain maturity that certainly helped me to achieve this PhD thesis in the best way. I am thankful for your availability, patience in explaining statistics and guidance in developing a critical mind. Being in your group, I learnt so much about the qualities a researcher should have. I enjoyed very much working in your group, be it as technical assistant or PhD student. After so many years in your group, it will be difficult to leave, and the end of this PhD is an important page that turns. Thank you also for accepting someone drinking Apfelshorle at social events. I won the challenge of achieving a PhD without coffee and bier! That is possible!

I would like to thank the members of my PhD advisory committee, Holger Goerlitz, Anne Charmantier and Ela Hau for their advice and guidance on the different projects. I would also like to thank the IMPRS for Organismal Biology for giving me the opportunity to participate to interesting courses and workshops. This was a great chance to further both my scientific and interpersonal skills. Thank you to IMPRS coordinators, Maggi Hieber, and Francisca Rosa Mende and Corrina Loës for their hard work and support.

My thanks go to my PhD fellow from the start, Alexander Hutfluss, for sharing the hurdles of our projects. Knowing someone else encounters the same difficulties relieves from thinking we may not be capable and helps moving forward. Thank you also for helping organising field work. My thanks also go to a former PhD student in the group, Maria Moiron, for sharing the long car rides to the LMU during which you could update me on all gossips. Many thanks for

your encouragements and your good mood. I thank Petri Niemelä for stimulating discussions and statistical advice.

I thank all members of the Behavioural Ecology group for the good atmosphere they bring to the group, for sharing time out of work around a beer (or Apfelshorle!) or for having dinner. These moments of fun are important to keep our well-being during a PhD.

Finally, I am grateful to my family, my parents and my brother, for their support all along these four years. Merci de m'avoir encouragé tout au long et de ne pas avoir douté que je pouvais accomplir un tel projet. Maman, papa, merci de m'avoir approvisionnée en produits du terroir français, et surtout en bon chocolat (!) et de m'avoir rendu visite de façon inconditionnelle chaque année. Et bien sûr, maman, merci pour les cures de rire à chaque visite qui sont des moyens de décompression très efficaces ! Je sais que ce n'était pas facile de me savoir si loin, mais je vous ai au moins permis de découvrir la belle région de Munich. Je vous suis reconnaissante de m'avoir fait confiance et de m'avoir laissé choisir mon chemin. Maman, merci aussi pour ces superbes aquarelles qui embellissent magnifiquement cette thèse. Enfin, je remercie Raphaël pour avoir été à mes côtés pendant la période la plus difficile de la thèse, la dernière année, endurante et avec les doutes de franchir les derniers obstacles. Merci pour tes conseils et les discussions, et surtout tes encouragements et ton réconfort.

Herzlichen Dank!

Author Contributions

Chapter 1

The study was conceived by N.J.D. and developed and operationalized with input from all authors. Analytical strategies designed, and statistical analyses performed by N.J.D., with input from M.M., and Y.G.A-A, and **A.M. A.M.** also coordinated the fieldwork and managed the database; all authors contributed to data collection. N.J.D. drafted the manuscript with input from all authors.

Chapter 2

Research was designed by **A.M.** and N.J.D., with input from all authors. **A.M.** and N.J.D. performed research. All authors collected data. E.M., M.N., J.L.Q., A.M.R., J.M.T., and K.v.O. provided data. **A.M.** assembled the dataset and analysed data. **A.M.** and N.J.D. wrote the paper. All authors provided input on manuscript.

Chapter 3

Research was designed and performed by **A.M.** and N.J.D. **A.M.** performed the experiment and collected data. **A.M.** analysed data and wrote the manuscript, with input from N.J.D.

Chapter 4

Research was designed and performed by **A.M.** and N.J.D. **A.M.** and N.J.D. collected data. **A.M.** analysed data and wrote the paper, with input from N.J.D.

Addresses of co-authors

Robin Abbey-Lee
IFM Biology, AVIAN Behavioural Genomics and Physiology Group, Linköping University, Linköping, Sweden; Email: robab01@liu.se

Yimen Araya-Ajoy
Center for Biodiversity Dynamics, Norwegian University of Science and Technology, Trondheim 7012, Norway; Email: yimen.araya-ajoy@ntnu.no

Ella F. Cole
Edward Grey Institute, Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom; Email: eleanor.cole@zoo.ox.ac.uk

Niels Dingemanse
Behavioural Ecology, Department of Biology II, Ludwig Maximilian University of Munich, 82152 Planegg-Martinsried

Erik Matthysen
Evolutionary Ecology Group, Department of Biology, University of Antwerp, B-2610 Antwerp, Belgium; Email: erik.matthysen@uantwerpen.be

Maria Moiron
Institute of Avian Research, 26386 Wilhelmshaven, Germany; Email: mariamoironc@gmail.com

Marion Nicolaus
Conservation Ecology Group, Groningen Institute for Evolutionary Life Science, University of Groningen, 9700 AB Groningen, The Netherlands; Email: nicolaus.marion@gmail.com

John L. Quinn
School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland T23 N73K; Email: J.Quinn@ucc.ie

Allison M. Roth
Department of Biology, University of Florida, Gainesville, FL 32611; Email: amr2264@columbia.edu

Joost M. Tinbergen
Conservation Ecology Group, Groningen Institute for Evolutionary Life Science, University of Groningen, 9700 AB Groningen, The Netherlands; Email: j.m.tinbergen@rug.nl

Kees van Oers
Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), 6708 PB Wageningen, The Netherlands; Email: K.vanOers@nioo.knaw.nl

Thijs van Overveld
Department of Conservation Biology, Estación Biológica de Doñana (CSIC), Américo Vespucio s/n E-41092 Sevilla, Spain; Email: mathijs.van.overveld@gmail.com

Statutory declaration and statement

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den 02.08.2021 Alexia Mouchet

Erklärung

Hiermit erkläre ich, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist, und ich mich anderweitig einer Doktorprüfung ohne Erfolg nicht unterzogen habe. Im Weiteren erkläre ich, dass ich mich mit Erfolg der Doktorprüfung im Hauptfach Biologie bei der Fakultät für Biologie der Ludwig-Maximilians-Universität München unterzogen habe.

München, den 02.08.2021 Alexia Mouchet